AD_____

Award Number: DAMD17-02-1-0057

TITLE: Mutagen Sensitivity, Apoptosis, and Polymorphism in DNA Repair as Measures of Prostate Cancer Risk

PRINCIPAL INVESTIGATOR: Radoslav Goldman, Ph.D.

CONTRACTING ORGANIZATION: Georgetown University
Washington, D.C. 20057

REPORT DATE: February 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-02-2006	Final	15 Jan 2002 – 14 Jan 2006

4. TITLE AND SUBTITLE

Mutagen Sensitivity, Apoptosis, and Polymorphism in DNA Repair as Measures of Prostate Cancer Risk

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-02-1-0057

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Radoslav Goldman, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Georgetown University
Washington, D.C. 20057

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
Original contains colored plates: ALL DTIC reproductions will be in black and white.

14. ABSTRACT

Prostate cancer is the most common lethal tumor among US males but etiology of the disease remains unknown. We hypothesize that low DNA repair contributes to increased risk of having prostate cancer. To evaluate the hypothesis, we conducted a case-control study of prostate cancer evaluating association of mutagen sensitivity phenotype with cancer risk. We established a repository of fully annotated specimen of 63 prostate cancer cases and 109 controls frequency matched on age and race. We created a sample repository consisting of serum, plasma, buffy coat, urine, toenail clipping and saliva. We also created a computerized database of the samples in Microsoft Access. We developed assays for mutagen sensitivity, comet assay, and apoptosis in white blood cells exposed to bleomycin and ionizing radiation to evaluate DNA repair capacity. We evaluated mutagen sensitivity in 95 subjects and determined that mean breaks in lymphocytes exposed to bleomycin are significantly higher ($p<0.001$) in prostate cancer cases (mean=1.1; SD=0.3) than controls (mean=0.7; SD=0.3).This pilot study fills important gaps in our understanding of prostate cancer etiology and produces new hypotheses which can be tested in an expanded prostate cancer study.

15. SUBJECT TERMS
DNA repair, prostate cancer, mutagen sensitivity, comet assay, single nucleotide polymorphism, molecular epidemiology

16. SECURITY CLASSIFICATION OF:				18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	150	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Introduction: Despite the fact that prostate cancer is the most common tumor among US males, relatively little is known about the causative mechanisms. The known risk factors include age, ethnicity or race, high-fat diet and family history of prostate cancer, but these factors are not sufficient for identification of men with increased susceptibility. This case-control study tested variation in the response to genotoxic stress as a biomarker of prostate cancer risk.

Mutagen sensitivity is an established biomarker of risk (1). Comet assay is an increasingly popular tool for human biomonitoring (2) with the potential to identify cancer-prone individuals in the general population (1). Both comet assay and mutagen sensitivity measure DNA damage in short-term cultured human lymphocytes exposed to bleomycin (or other mutagens) as either tail moment (comet assay) or number of chromatid breaks (mutagen sensitivity). While mutagen sensitivity is an established tool in population-based studies of cancer risk and was associated with increased risk of glioma, lung, colon, hepatocellular, and HN carcinoma, comet assay was used only recently in three pilot studies of breast, cervical, and lung cancer (1). Surprisingly, neither assay was used to study prostate cancer risk. Even though the exact mechanism underlying these phenotype is unknown, variability in DNA-repair capacity is consistent with the available experimental results (3). Moreover, it was shown in twin studies that mutagen sensitivity is heritable in non-cancer subjects. The correlation coefficient was 0.79 (95% confidence interval = 0.65-0.88) in monozygotic twins while for dizygotic twins the coefficient was 0.42 (95% confidence interval = 0.00-0.71) (4). Mutagen sensitivity phenotype therefore reflects multiple genetic traits related to DNA repair capacities, which predispose an individual to cancer risk. Comet assay has several advantages compared to mutagen sensitivity: 1. An independent measure of DNA repair; 2. Higher throughput and lower cost per assay; and 3. Smaller sample size (also called SCGE, single cell gel electrophoresis assay) (2). We examined the use of comet assay and mutagen sensitivity for screening of prostate cancer susceptibility.

Apoptosis is a molecular pathway eliminating, besides other functions, cells unable to cope efficiently with genotoxic stress. Deficient apoptosis is a likely candidate for a cancer-prone phenotype. Apoptosis was implicated in regulation of response to radiation therapy in prostate cancer (5), malignancy of prostatic tumor (6), and recurrence of prostate carcinoma following surgery (7). For example, in 54 prostate cancer patients treated with radiotherapy the response was negative in 84% cases with positive bcl-2 immunohistochemistry and bcl-2 was an independent prognostic variable for treatment with odds ratio of 7.3 (5). Apoptotic index was associated with disease recurrence in a study of 47 men following radical prostatectomy (7). Since the apoptotic phenotype is a composite measure of a number of converging mechanistic pathways, it may be advantageous to the measurement of each individual genotype in the pathway. Apoptosis was examined as a phenotypic predictor of prostate cancer risk in this study.

DNA repair consists of two major categories, excision repair (base excision repair and nucleotide excision repair) and recombination repair (homologous and non-homologous) (8). Numerous polymorphisms in the DNA repair genes have been identified (9) and are likely to contribute to cancer risk through decreased efficiency of response to genotoxic stress. But two functional polymorphisms in DNA repair genes, *OGG1* and *XRCC1*, are particularly relevant to this study. Both genes are involved in the repair of 8-hydroxy-guanine (8-OHdG) and other oxidative lesions (10); and our study

examines mainly how variability in the response to oxidative DNA damage modifies risk for prostate cancer (bleomycin is a radiomimetic which induces oxidative DNA damage and mutagen sensitivity is mainly a model of this pathway). OGG1 is a DNA glycosylase/AP lyase involved in base excision repair of 8-OHdG and XRCC1 is a DNA ligase III terminating the base excision repair cascade (10). The OGG1 Ser(321)Cys polymorphism codes for a protein with a lower 8-OHdG repair capacity and leads to several splicing variants of unknown functional significance (11). This variant occurs at a frequency of 0.4 in Japanese and was associated with an increased risk of lung cancer in a study of 241 cases and 197 controls with an OR=3.01 (95% CI 1.33-6.83) (12). This variant was found in a Caucasian population at a frequency of 0.22 and was not associated with lung cancer in this study (13). Examination of this polymorphism in prostate cancer is therefore highly relevant. The XRCC1 Arg(399)Gln polymorphism was associated with increased sensitivity of human lymphocytes to DNA damage (14), increased risk of squamous cell carcinoma of the head and neck (15), increased risk of early onset colorectal carcinoma (16), and increased risk of adenocarcinoma of the lung (17). The polymorphism occurs in 37% of Caucasians and 17% of African-Americans (19).

Body: This is a case-control study of phenotypic measures of prostate cancer risk. The recruitment was originally to be carried out by Dr. Trock (the recruitment was not budgeted as part of this grant) and phenotypic assay by the laboratory of Dr. Goldman as described in the proposal. Dr. Trock relocated to Johns Hopkins University, Baltimore, MD in 2002 and was not able to provide us with the needed blood samples. Because our assays have to be carried out on fresh blood, we were not able to find an alternative source of samples. To carry out the proposed research, we organized the recruitment of participants at Georgetown University. The recruitment of 100 cases and 100 controls proved to be a substantial task. We took advantage of additional funding of Dr. Goldman to accomplish the recruitment of participants as part of a larger study protocol called "Molecular Epidemiology of Prostate Cancer". The recruitment was carried out at the Georgetown University Hospital (GUH); we plan additional recruitment at the Veterans Administration Hospital (VAH), Washington, DC. This will allow us to recruit a sufficient number of African American participants for a comparison of DNA repair differences as a possible cause of the health disparity observed in prostate cancer in an expansion of the current study.

The recruitment of prostate cancer cases and matched controls at GUH and VAH was approved by the joint Medstar Research Institute-Georgetown University IRB and subsequently by the US Army Medical Research and Materiel Command's Human Research Protection Office (HRPO) (see appendix). The participants of this study are adult residents of the Washington, DC area, ages 18 and older. We enrolled all eligible patients prior to treatment that covered the full spectrum of tumor stage and grade. All subjects were briefly informed about the study by the attending physician and referred to a study coordinator. The study coordinator obtained informed consent, questionnaire data, and collected 45cc blood sample and urine/mouthwash. We developed the recruitment strategy in collaboration with our colleagues from the Department of Urology (Dr. Lynch), Radiation Oncology (Dr. Dritschilo), and Medical Oncology (Dr. Amin).

Cases were recruited among newly diagnosed prostate cancer patients at the above three departments.

Controls were recruited among visitors accompanying other patients to GUH (friends and spouses of patients in the Lombardi Cancer Center). We considered several methods of recruiting controls according to the described control selection guidelines. Random-digit phone dialing is likely to have low participation rates because we obtain blood sample for each participant; sibling controls could lead to overmatching on genetic factors; nominated peer controls were not an efficient group - most patients refuse to have their neighbors contacted because they do not want to disclose their disease state. We chose therefore visitors accompanying other patients to GU hospital. These controls are unbiased with respect to geography and socioeconomic status as they came to the hospital from the same referral area as the cancer cases. The subjects usually accompany a person to the hospital repeatedly, are motivated to participate, are easily contacted as they wait in the clinic, and typically do not make a special trip to the clinic for the study. It is a nonrandom subset, but was shown to be an excellent comparison group in several large studies.

The slow growth of prostate cancer and presence of asymptomatic cancer cases in the population presents a challenge to research studies. We used total serum PSA for all recruited controls to limit the possibility of including undetected cancer patients in the control group. We considered serum PSA>4.0 ng/ml and/or >2.5 ng/ml as uncertain, in agreement with the latest research (18). Assays were conducted on the larger sample; most restrictive analyses exclude all controls with PSA > 2.5 ng/ml. Inclusion of the PSA screening as part of the control selection protocol provides us with the opportunity to explain PSA testing and promote awareness of cancer screening. All controls with PSA > 2.5 ng/ml were given referrals to an urologist.

Cases were matched to controls on age (5 years) and race. It is important to match on these factors so that hypothesis testing is not compromised by severe imbalances in subject characteristics. We used frequency-matching whereby the proportions of cases and controls in each 5-year age group within each race category was held as closely similar as possible. In practice, this was accomplished by tabulating patient frequencies. This table showed the categories of race and age that were underrepresented among previously recruited controls, which helped the interviewer to choose an appropriate control.

We developed a research questionnaire inquiring about demographic information, reproductive history, tobacco use, alcohol consumption, general medical history and family history, occupational exposures, residential history, exercise, and education (see appendix) in collaboration with Dr. Ann Hsing, NCI. At a later stage, we added also a diet history questionnaire used by our colleagues at NCI (Dr. Ann Hsing). Every newly completed questionnaire was inspected for errors or inconsistencies prior to data entry. Double data entry was performed with automated range and consistency checks (in Microsoft Access). The files were protected by passwords and encryption.

Collection of biological specimen included blood, mouthwash, urine and toenail clipping. Each subject provided a 45 cc blood sample drawn into pre-labeled vacutainer glass tubes. Urine, toenail, and saliva were collected according to standard procedures. The samples were processed within 6 hours of collection at the GCRC core facility at Georgetown University. Case-control status was masked to the lab personnel since the

blood collection tubes show only a numeric study ID and a sample collection date. One green top tube was transferred to Dr. Goldman's laboratory for the mutagen sensitivity assay. Remaining blood samples were centrifuged and the blood components were separated into serum, clot, buffy coat, and plasma. The blood components were divided into aliquots of ~1 ml, bar coded, frozen at -80°C, and stored in a centrally monitored freezer facility. These samples constitute a repository of samples for prostate cancer research and serve for the testing of mutagen sensitivity and other endpoints as described below. In addition, we attempted to collect prostate tissue from radical prostatectomies but abandoned the project due to heterogeneity of the cancer tissue that requires assistance of a dedicated pathologist with tissue procurement. The ambitious sample collection plan was finally reduced in scope and we stopped collecting toenail clippings and mouthwash; currently, we collect mouthwash only if a blood sample cannot be obtained as reflected in **Table 1**.

Table 1	Collected						
	Enrolled	Blood	Urine	Mouthwash	Toenails	Interview	DHQ
Cases	72	69	64	42	42	62	33
Controls	117	114	110	75	73	108	63

Table 1. Summary of samples collected in course of our study. Diet history questionnaire (DHQ) was added later; collection of toenails and mouthwash was discontinued.

Our preliminary research indicated that Georgetown University sees about 150 new prostate cancer cases per year. Because of lack of funding and adequate personnel, we were not able to cover efficiently all the clinics seeing prostate cancer patients (Urology, Radiation Oncology, Medical Oncology) and controls. In addition, a portion of the 150 patients with newly diagnosed prostate cancer comes for a consultation, does not obtain treatment at GUH and does not wish to participate in our research study. We were able to recruit 72 prostate cancer cases and 117 controls frequency matched on age and race. Our effort which led to the establishment of a repository with fully annotated set of samples (blood, urine, and questionnaire) for 63 cases and 109 controls. The samples were bar coded with numerical identifiers, entered into a Microsoft Access based database, and stored at -80oC. The characteristics of the population are summarized in **Table 2**.

Table 2.	Cases n=63	Controls n=109
	(%)	(%)
Age		
less than 60	26	23
60 – 70	51	56
over 70	23	21
Race		
White	84	88
Black	16	12
Gleason Score		
<= 6	61	na
7-10	39	na
PSA		
<=4	14	85
>4	86	15

Table 2. All participants enrolled in our study.

Aim 1. Determine whether high mutagen sensitivity is associated with high prostate cancer risk.

For each person, a 62 hour culture of fresh whole blood collected in a green top (sodium heparin) vacutainer tube was established and the lymphocytes were stimulated with phytohemagglutinine (PHA). Following culture, the cells were exposed for 5 hours to bleomycin, fixed, and microscopic slides with chromosomal spreads were stained with Giemsa stain as described previously (20). All cases treated with antibiotics were excluded from consideration as antibiotics affect chromosomal breaks in the cultured lymphocytes. Mutagen sensitivity assay was completed on 35 cases and 60 controls. The remaining participants were not completed because of the following reasons: 1. Patients were not eligible for the mutagen sensitivity assay because of use of antibiotics or current infections (HIV, hepatitis); 2. The blood sample was missing of we could not resolve a scheduling conflict; or 3. Blood culture failed (contamination, handling errors). Majority of cases were not analyzed because of the use of antibiotics, especially cases recruited around the time of a biopsy procedure. Additional samples were not analyzed because we could not evaluate cultures with insufficient number of metaphases or condensed chromosomes. In the end, we compared results for 29 prostate cancer cases and 56 controls; all cases were newly diagnosed and enrolled prior to treatment. Description of the population analyzed for mutagen sensitivity is summarized in **Table 3**.

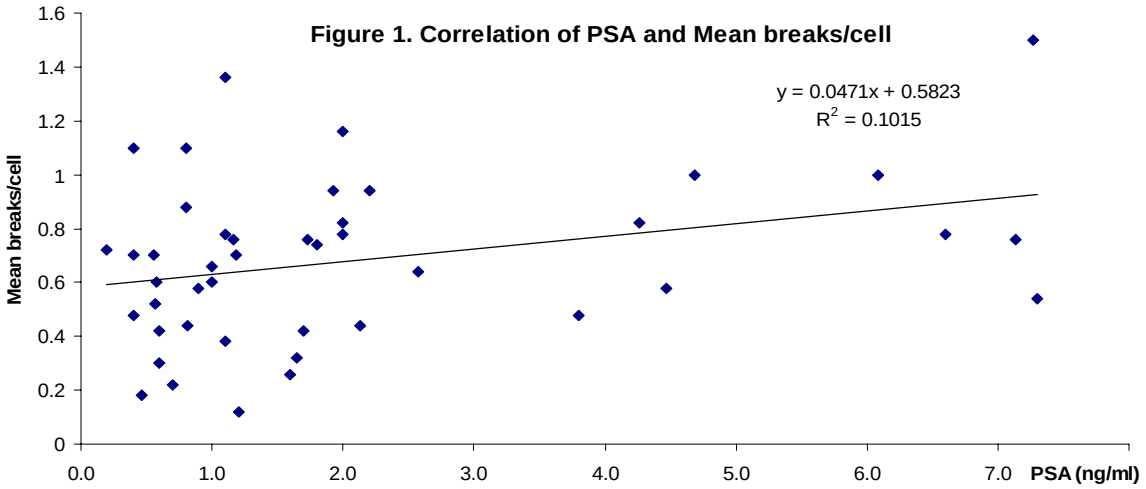
Table 3.	Cases n=35	Controls n=60
	(%)	(%)
Age		
less than 60	25	27
60 – 70	53	55
over 70	22	18
Race		
White	83	89
Black	17	11
Gleason Score		
<= 6	58	na
7-10	42	na
PSA (ng/ml)		
<=4	11	83
>4	89	17

Table 3. Participants with completed mutagen sensitivity assay

The results show that mean breaks in cases (mean=1.1, SD=0.3) are significantly higher ($p<0.001$) than in controls (mean=0.7, SD=0.3) **(Table 4)**. This is in agreement with our study hypothesis; however, the number of participants examined is smaller than anticipated due to the complications with patient source and recruitment. We are trying to increase the number of analyzed patients to confirm the finding and to summarize the results in a publication. It is interesting to note that our parallel study of mutagen sensitivity in head and neck cancer did not find significant difference in mutagen sensitivity (cases: n=39, mean=0.89, SD=0.4; controls n=54, mean=0.82; SD=0.3).

Table 4. Mean breaks	Mean	Median	Range		St dev
Cases (n=29)	1.1	1.0	0.5	1.7	0.3
Controls All (n=56)	0.7	0.7	0.1	1.8	0.3
Wout PSA outliers (n=48)	0.7	0.7	0.1	1.8	0.4

To further verify cancer free status in controls we performed a prostate specific antigen (PSA) test. Ten of the 60 controls had PSA > 2.5ng/ml (2.6-7.0 ng/ml). When we excluded this group of controls from the comparison of cases and controls for mutagen sensitivity, the results did not change (see Table 4). This suggests that patients with marginal elevation of PSA do not have increased mutagen sensitivity; this is not surprising given the weak association of marginal PSA (less than 10 ng/ml) with prostate cancer risk. **Figure 1** correlates PSA with mutagen sensitivity in this study population.



In addition to the mutagen sensitivity, we began evaluating comet assay as an alternative protocol for measurement of DNA damage/repair. This assay is an increasingly popular tool for human biomonitoring (1) with the potential to identify cancer-prone individuals in the general population (2). Both comet assay and mutagen sensitivity measure DNA damage in short-term cultured human lymphocytes exposed to bleomycin (or other mutagens). While mutagen sensitivity is an established tool in population-based studies of cancer risk and was associated with increased risk of glioma, lung, colon, hepatocellular, and HN carcinoma (1), comet assay was used only recently in three pilot studies of breast, cervical, and lung cancer (1). The largest of the studies examined 100 lung cancer patients and 110 controls using comet assay and found correlation of cancer risk with increased DNA damage (OR 4.2; CI 2.2-7.4) (21). In addition, DNA repair (measured as rate of damage disappearance) was an independent predictor of risk (OR 2.1; CI 1.1-4.0).

Our preliminary results are encouraging. We are finishing a paper describing optimization of the assay for testing of patient blood samples. A poster summarizing the results was presented at the 97th Annual AACR Conference, Washington, DC, April 2006. Our experiments follow published experimental settings with minor modifications as described below (21).

1) Coat microscope slide with normal melting point agarose (NMPA), solidify on ice for 5 min
2) Add cell suspension to low melting point agarose (LMPA) and form a layer of cell suspension on the NMPA coated slide
3) Dip the preparation in cold alkaline (pH>13) lysing solution (4^{o}C) for 3 hours

4) Transfer the preparations from lysing solution to alkaline electrophoresis buffer for 40 minutes to unwind DNA
5) Separate DNA for 25 minutes at 4^{o}C by alkaline electrophoresis using 0.92 V/cm and 300 mA current
6) Fix preparations with methanol, wash with distilled water
7) Stain with 0.01% ethidium bromide
8) Acquire 50 cell images per experiment (2 slides per experiment) using a fluorescent microscope with CDD camera (Olympus) and evaluate average fluorescent intensity in the head (intact nuclear DNA) and tail (damaged DNA) using comet imaging software (Loats Inc., Gaithersburg, MD). This imaging system was purchased by Lombardi Cancer Center and installed in our laboratory.

However, we incorporated the following important modifications: 1. Experiments were performed on whole blood stored prior to experiments at 4^{o}C overnight (as opposed to cultured lymphocyte). This standardizes and facilitates the handling of patient samples; 2. Whole blood was embedded in agarose prior to exposure. This facilitates measurement of repair kinetic; 3. We focused on measurement of DNA repair over 45 minutes following exposure to ionizing radiation. We will examine in future studies how DNA repair kinetic measured by comet assay correlates with cancer risk.

The above experimental setup was selected based on extensive testing of Jurkat T cells, short term cultured lymphocytes, and whole blood as briefly described below. A publication summarizing these results is currently in preparation. Lymphocytes from short term culture in the presence of PHA (62 hours) and IL2 (24 hours) were treated with 60 μg/ml bleomycin solution. Control samples were treated with the same volume of medium. After 30 min the samples were washed with fresh medium and subjected immediately to alkaline lysis (analysis of DNA damage) or incubated in fresh medium for 8 and 15 min at 37^{o}C before alkaline lysis (analysis of DNA damage repair). The experiment was done on three independent cultures from the same blood sample and each performed in duplicate for a total of 6 measurements at each dose/time **(Table 5)**.

Table 5. Reproducibility of Bleomycin Induced Comets

Experiment	0 ug/ml	60ug/ml 0min	60ug/ml 8min	60ug/ml 15min
1	0.964	90.023	77.01101124	31.513
2	0.163	92.257	57.45676768	35.2921
3	2.53	82.6808	31.4653	22.4742
4	5.58	76.0992	46.7498	18.5163
5	1.2675	38.998	13.1558427	1.355384615
6	1.3635	33.717	24.8144	3.1712
Mean	1.98	68.96	41.78	18.72
SD	1.92	25.94	23.35	14.11

This experiment (and several subsequent repeats with modifications) revealed that the measurement is not sufficiently reproducible between cultures to allow screening of samples in a population study. This prompted us to test ionizing radiation, which is known to yield the best results in terms of dosing and reproducibility. This experiment was done initially using 0-2 Gy of radiation, but even the highest dose resulted in only

minor increase in % tail DNA. As we are interested in the quantification of DNA repair, this dose was not sufficient and we increased the dose to 5-15 Gy subsequently. We did also modify the electrophoretic conditions by increasing electrophoresis time to 40 minutes. With these conditions, we achieved better reproducibility of the experiments as exemplified by the presented exposure to 15 Gy **(Table 6)**.

Table 6. Reproducibility of IR induced Comets

Experiment	0Gy	15Gy 0min	15Gy 8min	15Gy 15min
1	2.78	54.35	12.51	8.89
2	3.33	46.28	18.41	8.38
3	0.90	67.21	48.77	36.74
4	0.36	55.61	14.82	5.16
5	1.75	47.32	13.37	6.42
6	1.93	64.47	14.64	5.64
Mean	1.84	55.87	20.42	11.87
SD	1.11	8.60	14.03	12.27

The mean and standard deviation are summarized in **Figure 2**.

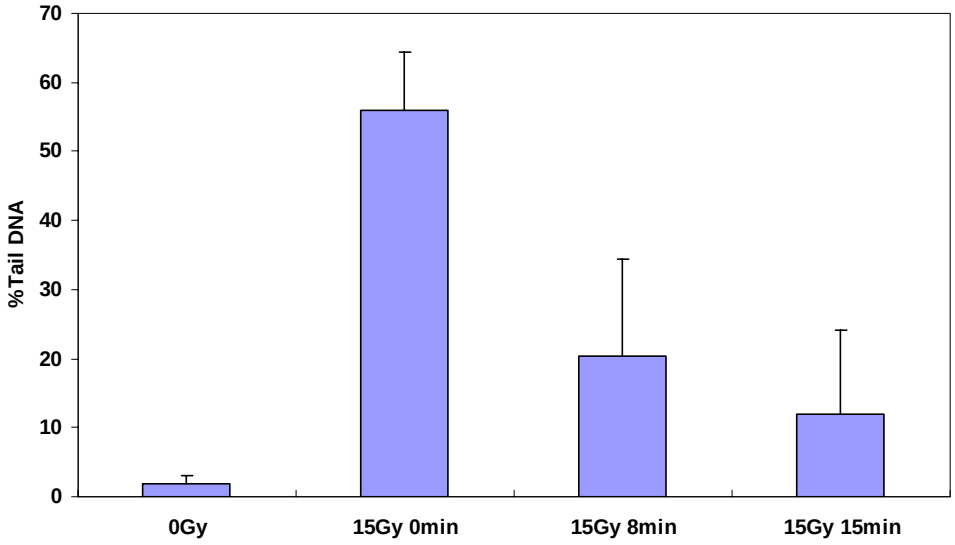


Figure 2. Kinetic of repair in cells exposed to 15 Gy of ionizing radiation.

We are investigating currently what percentage of cells undergoes apoptosis following the exposure to ionizing radiation, what is the kinetic of DNA repair at longer time points, and the reasons for the higher variability of the assays using bleomycin as the damaging agent. It was suggested in the literature that the repair of DNA damage following radiation is biphasic with a relatively fast repair of single strand breaks (within 15 minutes) and a slower repair of the residual damage, presumably double strand breaks, with a kinetic of hours. We hope to incorporate the optimized protocol into the

population study and compare the repair phenotypes measured by mutagen sensitivity and comet assay.

Upgrade of the fluorescent microscope and software for scoring of comets (LOATS Associates, Westminster, MD) and further adjustment of the experimental protocol resulted in use of lower doses and longer time points for DNA repair. Here we present comparison of dose response to 8 to 10 Gy of radiation and repair at 15 and 45 minute time point. The initial damage undergoes fast repair (within 15 minutes) and continues with a slow phase that is quantified at 45 minutes. We will examine patients under 9 Gy exposure at these time-points; we believe that all three time points provide independent information (damage, fast repair, and slow repair).

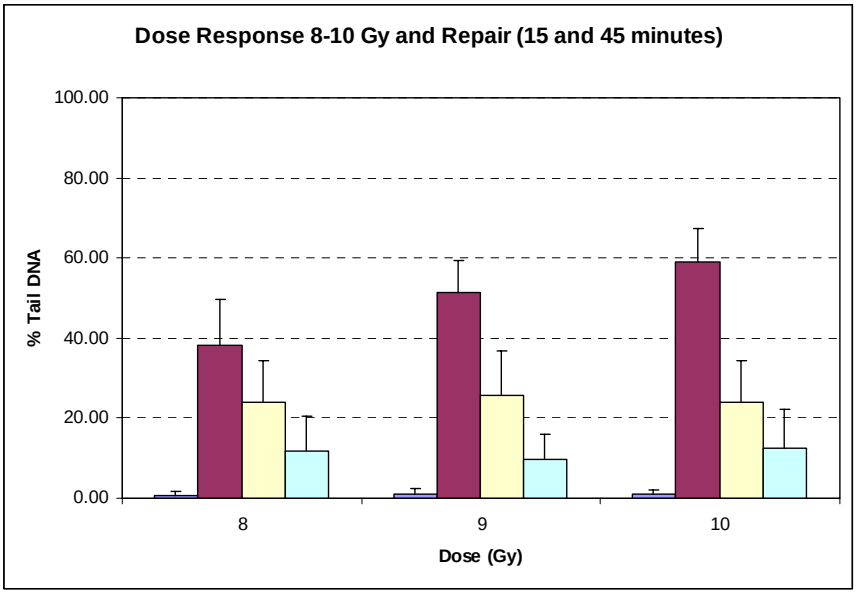


Figure 3:
Green series: unexposed
Red series: 0 min repair
Yellow series: 15 min repair
Blue series: 45 min repair

Initial comparison of repair kinetic in smoking (n=17) and non smoking (n=23) controls exposed to 9Gy IR showed that the repair kinetic in the interval between 0 and 15 minutes is faster in smokers **(Table 7)**.

Dose-Repair	0-0	9-0	9-15	9-45	Δ 9-0 to 9-15	Δ 9-15 to 9-45
	Smokers (n=17)					
Mean	1.01	47.07	23.58	14.42	23.49	9.15
SEM	0.11	2.12	2.00	1.35	1.55	1.10
	Ex-smokers (n=19) and Non-smokers (n=4)					
Mean	1.02	42.85	25.71	16.86	17.14	8.85
SEM	0.15	1.96	1.40	1.19	1.25	0.74
	T Test					
p-value	0.950	0.177	0.379	0.205	0.004	0.815

Table 7 (previous page). Comparison of repair kinetic in control smokers (n=17) and non (n=4) plus ex (n=19) smokers. Dose 0Gy and 9Gy, repair 0, 15, and 45 minutes.

The observed increase in fast repair kinetic in smokers suggests that DNA repair capacity for single strand break is induced by current exposure to cigarette smoke in peripheral blood lymphocytes.

Aim 2. Determine whether low apoptotic response is associated with increased prostate cancer risk.

We did perform Anexin V assay for phosphatidylserine flipping based on flow cytometry on 10 cases and 10 control samples of short term cultured lymphocytes **(Table 8)**.

Table 8. Apoptosis following exposure to bleomycin

	0ug/ml	20ug/ml	60ug/ml		0ug/ml	20ug/ml	60ug/ml
Control	12.92	31.38	38.74	case	10.13	26.32	39.76
Control	11.55	25.05	33.18	case	29.06	12.50	36.25
Control	9.04	19.90	30.20	case	35.01	44.96	47.22
Control	25.07	34.85	40.37	case	82.99	87.09	85.07
Control	11.58	29.77	40.29	case	20.82	37.72	39.32
Control	7.00	19.84	36.44	case	19.10	28.09	37.68
Control	11.96	25.04	37.54	case	18.50	25.67	41.42
Control	64.43	63.37	63.49	case	35.68	53.52	61.69
Control	33.96	44.19	49.92	case	36.39	50.83	53.98
Control	20.13	38.42	51.23	case	17.90	33.06	36.71
Mean	20.76	33.18	42.14	Mean	30.56	39.98	47.91
SD	17.43	13.18	9.96	SD	20.50	20.78	15.48

We did previously modify the tissue culture procedure by addition of IL2 following the culture in the presence of PHA in order to decrease variability of the assay. This worked reasonably well when performed on volunteer blood, but less well in the study as can be seen in Table 6. There are a number of samples with high background of Anexin V staining, especially in cancer cases. The reasons are at this point unknown. It is possible that the treatment with bleomycin is not sufficiently reproducible in this experimental setting even though we take care to use the same lot of reagent and aliquot the reagent as carefully as possible. We are currently evaluating the option to perform the apoptosis measurements on cells exposed to ionizing radiation and we are further optimizing the tissue culture protocol to eliminate the observed variability.

The exposure of lymphocytes to 0, 5, and 10 Gy of radiation led to small increase in apoptosis at 19 hour after exposure. We observe minimal effect of radiation immediately after exposure based on Anexin 5 staining. After 19 hours, percentage of cells in the first quadrant (FCS1) decreases with dose and quandrants 2 (FCS2, early apoptosis) and 3 (FCS3, late apoptosis) increase with dose. **Table 9** shows three individual experiments with mean and standard deviation.

Time (h)	Dose(Gy)	FCS1 (%)	FCS1 (%)	FCS1 (%)	**Mean1** **FCS1**	**Std. Dev1** **FCS1**
	0	77.17	89.66	95.20	**87.34**	**9.24**
0	5	74.86	90.04	94.66	**86.52**	**10.36**
	10	74.27	88.68	80.85	**81.27**	**7.21**
	0	78.36	83.72	73.98	**78.69**	**4.88**
19	5	70.60	75.20	54.22	**66.67**	**11.03**
	10	60.92	68.34	49.47	**59.58**	**9.51**
Time (h)	Dose(Gy)	FCS2 (%)	FCS2 (%)	FCS2 (%)	**Mean1** **FCS2**	**Std. Dev1** **FCS2**
	0	17.44	6.92	2.37	**8.91**	**7.73**
0	5	19.17	6.48	3.44	**9.70**	**8.34**
	10	19.84	7.22	12.60	**13.22**	**6.33**
	0	15.50	11.94	9.16	**12.20**	**3.18**
19	5	19.76	16.41	25.81	**20.66**	**4.76**
	10	26.16	21.06	29.15	**25.46**	**4.09**
Time (h)	Dose(Gy)	FCS3 (%)	FCS3 (%)	FCS3 (%)	**Mean1** **FCS3**	**Std. Dev1** **FCS3**
	0	4.05	3.09	1.28	**2.81**	**1.41**
0	5	5.10	3.17	1.86	**3.38**	**1.63**
	10	5.10	3.67	5.16	**4.64**	**0.84**
	0	4.98	3.82	16.59	**8.46**	**7.06**
19	5	8.04	7.42	19.52	**11.66**	**6.81**
	10	11.27	9.72	20.76	**13.92**	**5.98**

Aim 3. Determine whether the 'at risk' genetic variants of *OGG1* and *XRCC1* are risk factors for prostate cancer.

The testing of single nucleotide polymorphisms is not informative with respect to the small number of samples with results fro the phenotypic assays. Future expansion of the mutagen sensitivity and comet assay experiments should allow a meaningful comparison.

Key Research Accomplishments

1. The infrastructure for recruitment of cases and controls at Georgetown University Hospital was established. We developed recruitment strategies, research questionnaires, sample collection and storage protocols, databases for patient information (Epi Info and

Microsoft Access), and bar coded sample storage as documented in the appended materials.
2. We recruited 72 prostate cancer cases and 117 controls frequency matched on age and race. This allowed us to assemble a fully annotated repository of blood and urine sample of 63 prostate cancer cases and 109 controls matched on age and race.
3. The mutagen sensitivity assay was performed for 35 cases and 60 controls. The current result shows that mean breaks are significantly ($p<0.001$) higher in cases (mean=1.1, SD=0.3) than in controls (mean=0.7, SD=0.3).
4. We developed a complementary procedure for quantification of DNA repair capacity based on comet assay. This measurement was optimized to measure slow and fast repair kinetic at 9Gy exposure. This assay was tested on a pilot sample of controls. We compared DNA repair capacity in smokers and non-/ex-smokers which suggested that DNA repair with fast kinetic is induced in smokers.
5. Dr. Goldman (partial salary support on this grant) and his colleagues developed laboratory and computational methods for analysis of peptides associated with cancer. These methods (developed on a study of hepatocellular carcinoma) will be applied in the future to examination of prostate cancer patients.

Reportable Outcomes

The outcome of this research shows that mutagen sensitivity is higher in prostate cancer cases than controls. The completion of study was complicated by the lack of the planned source of research material (blood samples) caused by relocation of Dr. Trock from Georgetown University in 2002. While we developed alternative source of samples, part of the effort covering Dr. Goldman's salary was devoted to development of comet assay as an alternative method to assess cancer susceptibility and development of methods for analysis of peptides as diagnostic markers of cancer. The support from the current grant was acknowledged in several publication and posters published/presented recently by Dr. Goldman and colleagues. We plan to apply these methods to prostate cancer in future studies. A paper summarizing our experience with the comet assay is in preparation. We are trying to increase the size of the study of mutagen sensitivity to publish the results outlined in this report.

The PI, Radoslav Goldman and a Research Assistant, Xia Michelle Ma, were supported by the provided funding.

Papers:

1. Saha, D., Loeb, S., Wang, A., Pollock, A., and **Goldman, R**. (2006) Optimization of Comet Assay for Quantification of DNA Repair Capacity in Human Whole Blood. Mutation Research-Genetic Toxicology and Environmental Mutagenesis, *Submitted*
2. Ressom, H.W., Varghese, R.S., Orvisky, E., Drake, S.K., Hortin, G.L., Abdel-Hamid, M., Loffredo, C.A.; and **Goldman, R.** (2006) Ant Colony Optimization for Biomarker Identification from MALDI-TOF Mass Spectra. Proceedings of the IEEE EMBC Conference, New York, NY, *In Press*

3. Orvisky, E., Drake, S.K., Martin, B.M., Abdel-Hamid, M., Ressom, H., Varghese, R.S., An, Y., Saha, D., Hortin, G.L., Loffredo, C.A., **Goldman, R.** (2006) Enrichment of low molecular weight fraction of serum for mass spectrometric analysis of peptides associated with hepatocellular carcinoma. Proteomics, 6, 2895-2902.
4. Ressom, H., Varghese, R.S., Orvisky, E., Drake, S.K., Hortin, G.L., Loffredo, C.A., **Goldman, R.** (2005) Analysis of MALDI-TOF serum profiles for biomarker selection and sample classification. Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology La Jolla, CA, 378-384.
5. Ressom, H., Varghese, R.S., Abdel-Hamid, M., Eissa, S.A., Saha, D., Goldman, L., Petricoin, E.F., Conrads, T.P., Veenstra, T.D., Loffredo, C.A., **Goldman, R.** (2005) Analysis of mass spectral serum profiles for biomarker selection. Bioinformatics, 21, 4039-4045.
6. Ressom, H., Varghese, R.S., Saha, D., Orvisky, E., Goldman, L., Petricoin, E.F., Conrads, T.P., Veenstra, T.D., Abdel-Hamid, M., Loffredo, C.A., **Goldman, R.** (2005) Particle swarm optimization for analysis of mass spectral serum profiles. Genetic and Evolutionary Computation Conference, June 25-29, 2005, Washington, DC, H.-G. Beyer et al. (Editors), vol. 1, pp. 431-438.

Meeting Presentations:

1. Aleksandra Dakic, Allison Pollock, Michelle Ma, Daniel Saha, Sara Samie, Sherine Salem, Bozena Novotna, and **Radoslav Goldman**. Optimization of Comet assay for quantification of DNA repair capacity in human whole blood. 97th Annual AACR Conference, Washington, DC, April 2006
2. Yanming An, Mohamed Abdel-Hamid, Steve Drake, Lenka Goldman, Glen Hortin, Chris Loffredo, Eduard Orvisky, Habtom Ressom, Francoise Seillier-Moiseiwitsch, Rency Varghese, Antai Wang, and **Radoslav Goldman**. Analysis of serum peptides associated with hepatocellular carcinoma. 97th Annual AACR Conference, Washington, DC, April 2006
3. An, Y; Ressom, HW; Varghese, SA; Goldman, L; Orvisky, E; Liao, J; Wang, A; Seillier-Moiseiwitsch, F; Drake, SK; Hortin, GL; Loffredo, CA and **Goldman, R**. MALDI-TOF analysis of serum peptides associated with hepatocellular carcinoma. AACR Special Conference, New Developments in the Epidemiology of Cancer Prognosis: Traditional and Molecular Predictors of Treatment Response and Survival. Charleston, South Carolina, January 2006.
4. **Goldman, R**; An, Y; Liao, J; Orvisky, E; Ressom, HW; Varghese, SA; Goldman, L; Drake, SK; Hortin, GL; Loffredo, CA and Abdel-Hamid, M. MALDI-TOF analysis of serum peptides associated with hepatocellular carcinoma. ASPO, 30th Annual Meeting, Bethesda, MD, January 2006. Abstract of the oral presentation was published in *Cancer Epidemiol Biomarkers Prev* (2006) 15, 2, 409.
5. Ressom, H; Varghese, R; Dakic, A; Orvisky, E; Drake, SK; Hortin, GL; Abdel-Hamid, M; Loffredo, CA and **Goldman, R**. Analysis of MALDI-TOF Serum Profiles for Biomarker Selection and Sample Classification. American Association for the Study of Liver Diseases' (AASLD) Basic Research Single Topic Conference "Exploring the Functional Genomics and Proteomics of Liver in Health and Disease", Warrenton, VA, June 2005

6. Orvisky, E; Ressom, H; Wang, A; Saha, D; Goldman, L; Petricoin, EF; Conrads, TP; Veenstra, TD; Liotta, LA; Drake, SK; Hortin, GL; Abdel-Hamid, M; Loffredo, CA and **Goldman, R**. Enrichment of low molecular weight (LMW) serum fraction for MALDI-TOF detection of hepatocellular carcinoma (HCC). 96th Annual AACR Conference, Anaheim, CA, March 2005
7. Orvisky, E; Ressom, H; Saha, D; Goldman, L; Drake, SK; Hortin, GL; Abdel-Hamid, M; Loffredo, CA and **Goldman, R**. MALDI-TOF/TOF of enriched low molecular weight (LMW) serum fraction detects hepatocellular carcinoma (HCC). US Human Proteome Organization, 1st Annual Congress, Washington DC, March 2005
8. Orvisky, E; Ressom, H; Wang, A; Saha, D; Goldman, L; Drake, SK; Hortin, GL; Abdel-Hamid, M; Loffredo, CA and **Goldman, R**. Enrichment of low molecular weight (LMW) serum fraction for MALDI-TOF detection of hepatocellular carcinoma (HCC). Gordon Research Conference, New Frontiers in Cancer Detection and Diagnosis, Santa Ynez, CA, January 2005

Conclusions

The presented results suggest that mutagen sensitivity is higher in patients with prostate cancer compared to healthy controls. This may be related to lower DNA repair capacity but the sample size we examined is not sufficiently large for a definitive conclusion. The progress of the study was slowed down by the lack of the expected sample source. We have established alternative recruitment procedures, sample collection, processing, repository, and data management. We established a repository of fully annotated blood and urine samples of 63 prostate cancer patients and 109 controls frequency matched on age and race. This substantial effort was made possible by generous support from the Lombardi Cancer Center (GCRC, Biomarker Core, Histopathology and Tissue Core) and additional funding of Dr. Goldman. This resource and developed experimental methods will be utilized in subsequent studies.

References

1. Berwick, M. and Vineis, P. Markers of DNA Repair and Susceptibility to Cancer in Humans: an Epidemiologic Review. J Natl Cancer Inst. 6-7-2000;92(11):874-97.
2. Kassie, F., Parzefall, W., and Knasmuller, S. Single Cell Gel Electrophoresis Assay: a New Technique for Human Biomonitoring Studies. Mutat.Res 2000;463(1):13-31.
3. Wei, Q., Spitz, M. R., Gu, J., Cheng, L., Xu, X., Strom, S. S., Kripke, M. L., and Hsu, T. C. DNA Repair Capacity Correlates With Mutagen Sensitivity in Lymphoblastoid Cell Lines. Cancer Epidemiol Biomarkers Prev. 1996;5(3):199-204.
4. Cloos, J., Nieuwenhuis, E. J., Boomsma, D. I., Kuik, D. J., van der Sterre, M. L., Arwert, F., Snow, G. B., and Braakhuis, B. J. Inherited Susceptibility to Bleomycin-Induced Chromatid Breaks in Cultured Peripheral Blood Lymphocytes . J.Natl.Cancer Inst. 7-7-1999;91(13):1125-30.
5. Scherr, D. S., Vaughan, E. D., Wei, J., Chung, M., Felsen, D., Allbright, R., and Knudsen, B. S. BCL-2 and P53 Expression in Clinically Localized Prostate

Cancer Predicts Response to External Beam Radiotherapy. J Urol. 1999;162(1):12-6.
6. Lipponen, P. and Vesalainen, S. Expression of the Apoptosis Suppressing Protein Bcl-2 in Prostatic Adenocarcinoma Is Related to Tumor Malignancy. Prostate 6-15-1997;32(1):9-15.
7. Stapleton, A. M., Zbell, P., Kattan, M. W., Yang, G., Wheeler, T. M., Scardino, P. T., and Thompson, T. C. Assessment of the Biologic Markers P53, Ki-67, and Apoptotic Index As Predictive Indicators of Prostate Carcinoma Recurrence After Surgery. Cancer 1-1-1998;82(1):168-75.
8. Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. Human DNA Repair Genes. Science 2001;291:1284-1289.
9. Shen, M. R., Jones, I. M., and Mohrenweiser, H. Nonconservative Amino Acid Substitutions Exist at Polymorphic Frequency in DNA Repair Genes in Healthy Humans. Can. Res. 2-15-1998;58:604-8.
10. Boiteux, S.,and Radicella, J. P. The Human OGG1 Gene: Structure, Function, and its Implication in the Process of Carcinogenesis. Arch. Biochem. Biophys. 2000;377(1):1-8.
11. Kohno, T., Shinmura, K., Tosaka, M., Tani, M., Kim, S. R., Sugimura, H., Nohmi, T., Kasai, H., and Yokota, J. Genetic Polymorphisms and Alternative Splicing of the hOGG1 Gene that is Involved in Repair of 8-hydroxyguanine in Damaged DNA. Oncogene. 1998;16(25):3219-3225.
12. Sugimura, H., Kohno, T., Wakai, K., Nagura, K., Genka, K., Igarashi, H., Morris, B. J., and Yokota J. hOGG1 Ser326Cys Polymorphism and Lung Cancer Susceptibility. Cancer Epidemiol Biomarkers Prev. 1999;8(8):669-674.
13. Wikman, H., Risch, A., Klimek, F., Schmezer, P., Spiegelhalder, B., Dienemann, H., Kayser, K., Schulz, V., Drings, P., and Bartsch, H. hOGG1 Polymorphism and Loss of Heterozygosity (LOH): Significance for Lung Cancer Susceptibility in a Caucasian Population. Int J Cancer. 2000;88(6):932-937.
14. Duell, E. J., Wiencke, J. K., Cheng, T. J., Varkonyi, A., Zuo, Z. F., Ashok, T. D., Mark, E. J., Wain, J. C., Christiani, D. C., and Kelsey, K. T. Polymorphisms in the DNA Repair Genes XRCC1 and ERCC2 and Biomarkers of DNA Damage in Human Blood Mononuclear Cells [Published Erratum Appears in Carcinogenesis 2000 Jul;21(7):1457]. Carcinogenesis 2000;21(5):965-71.
15. Sturgis, E. M., Castillo, E. J., Li, L., Zheng, R., Eicher, S. A., Clayman, G. L., Strom, S. S., Spitz, M. R., and Wei, Q. Polymorphisms of DNA Repair Gene XRCC1 in Squamous Cell Carcinoma of the Head and Neck. Carcinogenesis 1999;20(11):2125-9.
16. Abdel-Rahman, S. Z., Soliman, A. S., Bondy, M. L., Omar, S., El Badawy, S. A., Khaled, H. M., Seifeldin, I. A., and Levin, B. Inheritance of the 194Trp and the 399Gln Variants of the DNA Repair Gene XRCC1 Are Associated With Colorectal Carcinoma in Egypt. Cancer Lett. 2000;159(1):79-86.
17. Divine, K. K., Gilliland, F. D., Crowell, R. E., Stidley, C. A., Bocklage, T. J., Cook, D. L., and Belinsky, S. A. The XRCC1 399 Glutamine Allele Is a Risk Factor for Adenocarcinoma of the Lung. Mutat.Res 1-5-2001;461(4):273-8.

18. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med 2004; 350(22):2239-2246.
19. Lunn, R. M., Langlois, R. G., Hsieh, L. L., Thompson, C. L., and Bell, D. A. XRCC1 Polymorphisms: Effects on Aflatoxin B1-DNA Adducts. Cancer Res. 6-1-1999;59(11):2557-61.
20. Zheng YL, Loffredo CA, Yu Z, Jones RT, Krasna MJ, Alberg AJ, Yung R, Perlmutter D, Enewold L, Harris CC, Shields PG. Bleomycin-induced chromosome breaks as a risk marker for lung cancer: a case-control study with population and hospital controls. Carcinogenesis. 2003 Feb;24(2):269-274.
21. Schmezer, P., Rajaee-Behbahani, N., Risch, A., Thiel, S., Rittgen, W., Drings, P., Dienemann, H., Kayser, K. W., Schulz, V., and Bartsch, H. Rapid Screening Assay for Mutagen Sensitivity and DNA Repair Capacity in Human Peripheral Blood Lymphocytes. Mutagenesis 2001;16(1):25-30.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

Informed Consent for Clinical Research (cases)

MedStar Research Institute/Georgetown Medical Center

INSTITUTION: GUMC + WHC

INTRODUCTION

We invite you to take part in a research study. The study is called 'Molecular Epidemiology of Prostate Cancer'. Please take your time to make your decision. Discuss it with your family and friends. It is important that you read and understand several general principles that apply to all who take part in our studies:

(a) Taking part in the study is entirely voluntary;

(b) Personal benefit to you may or may not result from taking part in the study, but knowledge may be gained from your participation that will benefit others;

(c) You may withdraw from the study at any time without any of the benefits you would have received normally being limited or taken away.

The nature of the study, the benefits, risks, discomforts and other information about the study is discussed below. Any new information discovered, at any place during the research, which might affect your decision to participate or remain in the study will be provided to you. You are urged to ask the staff members any questions you have about this study and the staff members will explain the questions to you. The investigator (person in charge of this research study) is Dr. Radoslav Goldman. The research is being sponsored by the Department of Defense. The Department of Defense is called the sponsor and the Georgetown University is being paid by the Department of Defense to conduct this study with Dr. Radoslav Goldman as the primary investigator.

WHY IS THE STUDY BEING DONE?

You are being asked to participate in this study because you are suspected of having prostate cancer or have prostate cancer. Your prostate tumor, blood and other samples may show us how cancer develops and what are the factors that helped increase the cancer risk.

The purpose of this study is to learn about the natural history of prostate cancer and its causes and treatments. This research is being done because the causes of prostate cancer are not well understood at present. The purpose of this research is to see how someone's ability to respond to genetic damage

modifies risk of prostate cancer. We will test how your ability to repair damaged DNA and eliminate cells that did not repair the damage modifies prostate cancer risk.

We will examine your blood, cheek samples, saliva, nail clippings and urine to see if tests for your response to chemical exposure can help us predict who might be at greater risk of prostate cancer. If you are going to have surgery, or had surgery, or if you are going to have a biopsy or had a biopsy, we will use samples of tumor tissue, as well as adjacent normal tissue, to determine whether markers in the tissue suggest how the cancer developed. The specimen will not be used for diagnostic purposes or for purposes related to your medical care. That is, the experiments done on these samples will not be used for decisions about your personal risk of prostate cancer, your treatment or your prognosis. These specimens will be available to qualified medical researchers for scientific studies that have been approved by the Principal Investigator, listed above, and an oversight committee. Researchers who receive these samples will not have access to your name or other identification information.

If you wish, you will be given the opportunity to identify friends living in your geographical area to be controls in the study. This would help us to identify a group of controls subjects without prostate cancer. We hope that this research can lead to the discovery of new tests for cancer risk, including genetic tests.

All men older than 18 years of age at all stages of presentation are eligible to participate in this study.

HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?

About 600 people (300 patients and 300 controls) will take part in this study and will be recruited at Washington Hospital Center and Georgetown University Medical Center. Participants in the study are referred to as "subjects".

WHAT IS INVOLVED IN THE STUDY?

Upon reviewing and signing this informed consent, you will begin the study. We will ask you questions using a form that will take about an hour to finish. If you do not want to do the whole questionnaire at the time you give blood, we can do only one part lasting about 15 minutes and then we will contact you later to finish the study. This research will be conducted on an experimental basis only, and you will not be provided with any information about your test results.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

If you take part in this study, you will have the following tests and procedures

1. Upon reviewing and signing this informed consent, you will begin the study.
2. Undergo an in person interview lasting about one hour administered by a trained interviewer.
3. Provide a blood sample that is about 3 tablespoons.
4. Provide a urine specimen.
5. Provide two cheek swab samples.
6. Provide saliva
7. Provide nail clipping.
8. Allow us to use the unneeded portion of your prostate tissue, as well as a small sample of adjacent normal tissue for research purposes.

HOW LONG WILL I BE IN THE STUDY?

We expect that your participation in the study will take an extra hour in addition to your scheduled examination. The study is completed after you finish your questionnaire and donate your blood, urine, nail, cheek sample, saliva and tissue from surgery/biopsy not needed for diagnostic purposes. However, if you agree below, we may call you in the future for additional information and/or sample collection. We will use your sample for different tests as described above and as new hypotheses develop for as long as it lasts and is useful for our testing. If the sample is no longer useful, it will be destroyed. However, you can request that your blood, cheek, saliva, nail, urine and prostate tissues be destroyed at any time. To have your samples destroyed, you can contact Dr. Goldman at 202-687 9868.

The investigators, physicians or sponsors may stop the study or take you out of the study at any time should they judge that it is in your best interest to do so, if you experience a study-related injury, or if you do not comply with the study plan. They may remove you from the study for various other administrative and medical reasons. They can do this without your consent.

In the future, it might be necessary to contact you for further information or an additional blood sample (or other type of biological sample). If this is okay, please indicate below. You can refuse to do so now or later. Please check and initial below:

I ____may ____may not be contacted in the future for further information or biological samples.

_____ Sign your initials here.

WHAT ARE THE RISKS OF THE STUDY?

There is a very slight chance of a bruise or an infection from the blood draw, but we use only trained medical technicians to draw your blood and they will use the best available precautions. Another possible risk is that your genetic information might be obtained by persons from outside the study. We will minimize this chance by maintaining the confidentiality of your test results and study records at all times (see below). For more information about risks and side effects, ask the research staff or contact Radoslav Goldman at 202-687- 9868.

ARE THERE ANY BENEFITS TO TAKING PART IN THE STUDY?

If you agree to take part in this study, there is no direct medical benefit to you. We hope the information learned from this study will benefit others in the future.

WHAT ABOUT CONFIDENTIALITY?

Efforts will be made to protect your personal information to the extent allowed by law. Medical records of research study participants are stored and kept according to legal requirements. You will not be identified in any reports or publications resulting from this study. Organizations that may request, inspect and/or copy your research and medical records for quality assurance and data analysis include groups such as: Department of Defense, Food and Drug Administration, MedStar Research Institute, Georgetown University, and Institutional Review Board (IRB).

We will store your tissue, blood, cheek, saliva, nail and urine samples, or genetic material prepared from your blood, urine, cheek, nail or prostate tissue, in a secure room with restricted access. Only people working on this research project can work on your sample. Because we want to protect your confidentiality, your samples will have only a number on the tube and will not have your name or other identifier information.

We will protect your genetic and other testing results. We will control access to the computer files that hold this information. Access to the computer files can only be obtained through multiple passwords. Only authorized study personnel can link your sample to you. This information will not be released to anyone. “Anyone” includes you, your family, your doctor, your insurance company, or your employer. This is because the research is at a very early stage and we would not be able to tell you what your results mean. This information will not be included in any medical records.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

CERTIFICATE OF CONFIDENTIALITY

To help us protect your privacy, we have obtained a Certificate of Confidentiality from the National Institutes of Health. With this Certificate, the researchers cannot be forced to disclose information that may identify you, even by a court subpoena, in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings. The researchers will use the Certificate to resist any demands for information that would identify you, except as explained below.

The Certificate cannot be used to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of Federally funded projects or for information that must be disclosed in order to meet the requirements of the federal Food and Drug Administration (FDA).

You should understand that the Certificate of Confidentiality does not prevent you or a member of your family from voluntarily releasing information about yourself or your involvement in this research. If an insurer, employer, or other person obtains your written consent to receive research information, then the researchers may not use the Certificate to withhold that information.

WHAT ARE THE COSTS?

There is no cost to participate in the study.

You should not expect any one to pay you for pain, worry, lost income, or non-medical care costs that occur from taking part in this research study.

You or your insurance company will be charged for continuing medical care and/or hospitalization that are not a part of the study.

RESEARCH RELATED INJURY

The Department of Defense is partially funding this research. Should you be injured as a direct result of participating in this research, you will be provided medical care at no cost to you. You will not receive any injury compensation, only medical care. Your insurance company will be billed, but you will not be liable for any costs not covered by your insurance. Additional information on this subject

may be obtained from the Office of the Medical Director, Georgetown University Hospital at (202) 784-3011.

You *will not* be paid for participating in this study.

COMMERCIAL INTEREST

On rare occasions, laboratory research on human specimens results in discoveries that are the basis for new research products or diagnostic and therapeutic methods. It is the policy of Georgetown University Medical Center, MedStar, Inc., and their affiliates not to compensate you for any future financial claim to your tissues for research and development for commercial and noncommercial purposes. No funds are available or will be paid by the MedStar Research Institute, MedStar Health or Georgetown University to repay you in case of injury.

_____ I understand that I will not receive financial compensation for my biological samples at any time. (sign initials here)

WHAT ARE MY RIGHTS AS A PARTICIPANT?

Taking part in this study is voluntary. You may choose not to take part in or leave the study at any time. If you request, the link between your name and the study results will be destroyed. Also, your biological samples will be discarded at your request. However, the results of any finished analysis and or published result will be kept to preserve the validity of the study. If you choose to not take part in or to leave the study, your regular care will not be affected and you will not lose any of the benefits you would have received normally.

We will not provide you with any of the results we obtain from your biological samples.

We will tell you about new information that may affect your health, welfare, or participation in this study.

WHO DO I CALL IF I HAVE QUESTIONS OR PROBLEMS?

For questions about the study, problems, unexpected physical or psychological discomforts or injuries related to the study, contact day or night the research doctor, Radoslav Goldman at 202-687 9868. If you would like to write to him, please send mail to: Radoslav Goldman,

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

Georgetown University, 3970 Reservoir Road NW, Research Building W309A, Washington DC 20057.

If you are a participant at Washington Hospital Center and have questions about your rights as a research participant, contact the MedStar Research Institute. Direct your questions to Dr. Barbara Howard at Medstar Research Institute:

MedStar Research Institute
6495 New Hampshire Ave., Suite 201
Hyattsville, MD 20783
Tel: (301) 853-7532
Pager: 1-888-663-6842

If you are a participant at Georgetown University Medical Center and have questions about your rights as a research participant, contact the Georgetown University IRB Office. Direct your questions to:

Ms. Laura Miller, Executive Officer, Institutional Review Board at:
Address: Georgetown University Medical Center Telephone: (202) 687-1506
3900 Reservoir Road, N.W.
NE 105 Med-Dent
Washington, D.C. 20007

SIGNATURES

As a representative of this study, I have explained the purpose, the procedures, the benefits and risks that are involved in this research study. Any questions that have been raised have been answered to the individuals satisfaction.

_____ _____

Signature of person obtaining the consent Date

I, the undersigned have been informed about this study's purpose, procedures, possible benefits and risks, and I have received a copy of this consent. I have been given the opportunity to ask questions before I sign, and I have been told that I can ask other questions at any time. I voluntarily agree to participate in this study. I am free to withdraw from the study at any time without need to justify my

decision. This withdrawal will not in any way effect my future treatment or medical management. I agree to cooperate with Dr. Radoslav Goldman and the research staff and to inform them immediately if I experience any unexpected or unusual symptoms.

_____

_____

Name and Permanent Address of Subject (Printed)

_____ _____

Signature of Subject Date

_____ _____

Signature of Witness Date

_____ _____

Principal Investigator (if not person obtaining consent) Date

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

Follow up Sample Acquisition Consent

As a continuation of the study in which I enrolled on _____ (date), I agree to provide a set of biological samples including urine, blood (about 3 tablespoons), cheek cells, and saliva and to answer questions about my medical history. In case I undergo surgery to remove a tumor, I agree to donate the unneeded portion of my head and neck tissue as well as adjacent normal tissue removed at surgery for research purposes. I, the undersigned, have been informed about this study's purpose, procedures, possible benefits and risks, and I have received a copy of this consent. I have been given the opportunity to ask questions before I sign, and I have been told that I can ask other questions at any time. I voluntarily agree to participate in this study. I am free to withdraw from the study at any time without need to justify my decision. This withdrawal will not in any way effect my future treatment or medical management. I agree to cooperate with Dr. Radoslav Goldman and the research staff and to inform them immediately if I experience any unexpected or unusual symptoms related to the research study.

_____ _____

Signature of Subject Date

_____ _____

Signature of Witness Date

_____ _____

Principal Investigator (if not person obtaining consent) Date

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

Informed Consent for Clinical Research (controls)

MedStar Research Institute/Georgetown Medical Center

INSTITUTION: GUMC + WHC

INTRODUCTION

We invite you to take part in a research study. The study is called 'Molecular Epidemiology of Prostate Cancer'. Please take your time to make your decision. Discuss it with your family and friends. It is important that you read and understand several general principles that apply to all who take part in our studies:

(a) Taking part in the study is entirely voluntary;

(b) Personal benefit to you may or may not result from taking part in the study, but knowledge may be gained from your participation that will benefit others;

(c) You may withdraw from the study at any time without any of the benefits you would have received normally being limited or taken away.

The nature of the study, the benefits, risks, discomforts and other information about the study is discussed below. Any new information discovered, at any place during the research, which might affect your decision to participate or remain in the study will be provided to you. You are urged to ask the staff members any questions you have about this study and the staff members will explain the questions to you. The investigator (person in charge of this research study) is Dr. Radoslav Goldman. The research is being sponsored by the Department of Defense. The Department of Defense is called the sponsor and the Georgetown University is being paid by the Department of Defense to conduct this study with Dr. Radoslav Goldman as the primary investigator.

WHY IS THE STUDY BEING DONE?

You are being asked to participate in this study because a comparison group free of prostate cancer is needed to evaluate the results. Your blood and other samples may show us how cancer develops and what the factors are that help increase cancer risk.

The purpose of this study is to learn about the natural history of prostate cancer and its causes and treatments. This research is being done because the causes of prostate cancer are not well understood at present. The purpose of this research is to see how someone's ability to respond to genetic damage

modifies risk of prostate cancer. We will test how your ability to repair damaged DNA and eliminate cells that did not repair the damage modifies prostate cancer risk.

We will examine your blood, cheek swabs, saliva, nail clippings and urine to see if tests for your response to chemical exposure can help us predict who might be at greater risk of prostate cancer. The specimens will not be used for diagnostic purposes or for purposes related to your medical care. That is, the experiments done on these samples will not be used for decisions about your personal risk of prostate cancer. These specimens will be available to qualified medical researchers for scientific studies that have been approved by the Principal Investigator, listed above, and an oversight committee. Researchers who receive these samples will not have access to your name or other identification information. We hope that this research can lead to the discovery of new tests for cancer risk, including genetic tests.

Men older than 18 years of age free of prostate cancer are eligible to participate in this study. To minimize the possibility that you have undetected prostate cancer, we will perform a test for prostate specific antigen (PSA) on a portion of your blood sample free of charge to you. If your test shows a PSA value greater than 2.5ng/ml, a follow up examination by a doctor will be recommended.

_____ (please initial) I agree to have my PSA level tested .

_____ (please initial) I agree to have my physician notified at the following address if the PSA level is elevated. If you do not have a physician, we recommend that you contact one in case the PSA level is elevated.

Physician's name: _____

Address:_____
_____
_____

Phone:_____ Fax:_____

HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?

About 600 people (300 patients and 300 controls) will take part in this study and will be recruited at Washington Hospital Center and Georgetown University Medical Center. Participants in the study are referred to as "subjects".

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

WHAT IS INVOLVED IN THE STUDY?

Upon reviewing and signing this informed consent, you will begin the study. We will ask you questions using a form that will take about an hour to finish. If you do not want to do the whole questionnaire at the time you give blood, we can do only one part lasting about 15 minutes and then we will contact you later to finish the study. Your blood, cheek cells, saliva, nail tissue, and urine will be tested for their response to chemical exposure, in order to identify tests that may predict cancer risk. This research will be conducted on an experimental basis only, and apart from your PSA test results, you will not be provided with any other information.

If you take part in this study, you will have the following tests and procedures:

1. Upon reviewing and signing this informed consent, you will begin the study.
2. Undergo an in person interview lasting about one hour administered by a trained interviewer.
3. Provide a blood sample that is about 3 tablespoons. One of the samples will be tested to determine your PSA level.
4. Provide a urine specimen.
5. Provide two cheek swab samples.
6. Provide saliva.
7. Provide nail clippings.

HOW LONG WILL I BE IN THE STUDY?

We expect that your participation in the study will take about an hour. The study is completed after you complete your questionnaire and donate your blood, urine, nail clippings, saliva and a cheek sample. However, if you agree below, we may call you in the future for additional information and/or sample collection. We will use your sample for different tests as described above and as new hypotheses develop for as long as it lasts and is useful for our testing. If the sample is no longer useful, it will be destroyed. However, you can request that your blood, cheek cells, saliva, nail tissue, and urine be destroyed at any time. To have your samples destroyed, you can contact Dr. Goldman at 202-687-9868.

The investigators, physicians or sponsors may stop the study or take you out of the study at any time should they judge that it is in your best interest to do so, if you experience a study-related injury, or if you do not comply with the study plan. They may remove you from the study for various other administrative and medical reasons. They can do this without your consent.

In the future, it might be necessary to contact you for further information or an additional blood sample (or other type of biological sample). If this is okay, please indicate below. You can refuse to do so now or later. Please check and initial below:

I ____may ____may not be contacted in the future for further information or biological samples.

_____ Sign your initials here.

WHAT ARE THE RISKS OF THE STUDY?

There is a very slight chance of a bruise or an infection from the blood draw, but we use only trained medical technicians to draw your blood and they will use the best available precautions. Another possible risk is that your genetic information might be obtained by persons outside the study. We will minimize this chance by maintaining the confidentiality of your test results and study records at all times (see below).
For more information about risks and side effects, ask the research staff or contact Radoslav Goldman at 202-687 9868.

ARE THERE ANY BENEFITS TO TAKING PART IN THE STUDY?

If you agree to take part in this study, there is no direct medical benefit to you. We hope the information learned from this study will benefit others in the future.

WHAT ABOUT CONFIDENTIALITY?

Efforts will be made to protect your personal information to the extent allowed by law. Medical records of research study participants are stored and kept according to legal requirements. You will not be identified in any reports or publications resulting from this study. Organizations that may request, inspect and/or copy your research and medical records for quality assurance and data analysis include groups such as: Department of Defense, Food and Drug Administration, MedStar Research Institute, Georgetown University, and Institutional Review Board (IRB).
We will store your blood, cheek, saliva, nail and urine samples, or genetic material prepared from your blood, urine, cheek, saliva and nail in a secure room with restricted access. Only people working on this research project can work on your samples. Because we want to protect your confidentiality, your samples will have only a number on the tube and will not have your name or other identifier information.

We will protect your genetic and other testing results. We will control access to the computer files that hold this information. Access to the computer files can only be obtained through multiple passwords. Only authorized study personnel can link your sample to you. This information will not be released to anyone. "Anyone" includes you, your family, your doctor, your insurance company, or your employer. This is because the research is at a very early stage and we would not be able to tell you what your results mean. This information will not be included in any medical records.

CERTIFICATE OF CONFIDENTIALITY

To help us protect your privacy, we have obtained a Certificate of Confidentiality from the National Institutes of Health. With this Certificate, the researchers cannot be forced to disclose information that may identify you, even by a court subpoena, in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings. The researchers will use the Certificate to resist any demands for information that would identify you, except as explained below.

The Certificate cannot be used to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of Federally funded projects or for information that must be disclosed in order to meet the requirements of the federal Food and Drug Administration (FDA).

You should understand that the Certificate of Confidentiality does not prevent you or a member of your family from voluntarily releasing information about yourself or your involvement in this research. If an insurer, employer, or other person obtains your written consent to receive research information, then the researchers may not use the Certificate to withhold that information.

WHAT ARE THE COSTS?

There is no cost to participate in the study
You should not expect any one to pay you for pain, worry, lost income, or non-medical care costs that occur from taking part in this research study.

You or your insurance company will be charged for continuing medical care and/or hospitalization that are not a part of the study.

RESEARCH RELATED INJURY

The Department of Defense is partially funding this research. Should you be injured as a direct result of participating in this research, you will be provided medical care at no cost to you. You will not receive any injury compensation, only medical care. Your insurance company will be billed, but you will not be liable for any costs not covered by your insurance. Additional information on this subject may be obtained from the Office of the Medical Director, Georgetown University Hospital at (202) 784-3011.

You *will not* be paid for participating in this study.

COMMERCIAL INTEREST

On rare occasions, laboratory research on human specimens results in discoveries that are the basis for new research products or diagnostic and therapeutic methods. It is the policy of Georgetown University Medical Center, MedStar, Inc., and their affiliates not to compensate you for any future financial claim to your tissues for research and development for commercial and noncommercial purposes. No funds are available or will be paid by the MedStar Research Institute, MedStar Health or Georgetown University to repay you in case of injury.
I understand that I will not receive financial compensation for my biological samples at any time.
_____(sign initials here)

WHAT ARE MY RIGHTS AS A PARTICIPANT?

Taking part in this study is voluntary. You may choose not to take part in or leave the study at any time. If you request, the link between your name and the study results will be destroyed. Also, your biological samples will be discarded at your request. However, the results of any finished analysis and or published result will be kept to preserve the validity of the study. If you choose to not take part in or to leave the study, your regular care will not be affected and you will not lose any of the benefits you would have received normally.
We will tell you about new information that may affect your health, welfare, or participation in this study.

WHO DO I CALL IF I HAVE QUESTIONS OR PROBLEMS?

For questions about the study, problems, unexpected physical or psychological discomforts or injuries related to the study, contact day or night the research doctor, Radoslav Goldman at 202-687-9868. If you would like to write to him, please send mail to: Radoslav Goldman, Georgetown University, 3970 Reservoir Road NW, Research Building W309A, Washington DC 20057.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

If you are a participant at Washington Hospital Center and have questions about your rights as a research participant, contact the MedStar Research Institute. Direct your questions to Dr. Barbara Howard at Medstar Research Institute:

MedStar Research Institute
6495 New Hampshire Ave., Suite 201
Hyattsville, MD 20783
Tel: (301) 853-7532
Pager: 1-888-663-6842

Or

If you are a participant at Georgetown University Medical Center and have questions about your rights as a research participant, contact the Georgetown University IRB Office. Direct your questions to:

Ms. Laura Miller, Executive Officer, Institutional Review Board at:
Address: Georgetown University Medical Center Telephone: (202) 687-1506
3900 Reservoir Road, N.W.
NE 105 Med-Dent
Washington, D.C. 20007

SIGNATURES

As a representative of this study, I have explained the purpose, the procedures, the benefits and risks that are involved in this research study. Any questions that have been raised have been answered to the individual's satisfaction.

_____ _____

Signature of person obtaining the consent Date

I, the undersigned have been informed about this study's purpose, procedures, possible benefits and risks, and I have received a copy of this consent. I have been given the opportunity to ask questions before I sign, and I have been told that I can ask other questions at any time. I voluntarily agree to participate in this study. I am free to withdraw from the study at any time without need to justify my decision. This withdrawal will not in any way effect my future treatment or medical management. I

agree to cooperate with Dr. Radoslav Goldman and the research staff and to inform them immediately if I experience any unexpected or unusual symptoms.

_____

_____

Printed name and permanent address of subject.

_____ _____

Signature of Subject Date

_____ _____

Signature of Witness Date

_____ _____

Principal Investigator (if not person obtaining consent) Date

Study number:_ Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

Follow up Sample Acquisition Consent

As a continuation of the study in which I enrolled on _____ (date), I agree to provide a set of biological samples including urine, blood (about 3 tablespoons), cheek cells, and saliva and to answer questions about my medical history. I, the undersigned, have been informed about this study's purpose, procedures, possible benefits and risks, and I have received a copy of this consent. I have been given the opportunity to ask questions before I sign, and I have been told that I can ask other questions at any time. I voluntarily agree to participate in this study. I am free to withdraw from the study at any time without need to justify my decision. This withdrawal will not in any way effect my future treatment or medical management. I agree to cooperate with Dr. Radoslav Goldman and the research staff and to inform them immediately if I experience any unexpected or unusual symptoms related to the research study.

_____ _____

Signature of Subject Date

_____ _____

Signature of Witness Date

_____ _____

Principal Investigator (if not person obtaining consent) Date

MedStar Research Institute-Georgetown University Oncology Institutional Review Board Application (Protocol) IRB Review (AB-1)

Section One: Application Information

Principal Investigator	Radoslav Goldman, Ph.D.
Department	Oncology
Title	Assistant Professor
Phone/Pager: 202-687 9868	**Fax:** 202-687 1988
E-mail address:rg26@georgetown.edu	
Mailing Address: Georgetown University, Lombardi Cancer Center, LL (S) Level, Room 183, 3800 Reservoir Rd. NW, Washington DC 20057	
Co-Investigator: Christopher Loffredo, Department of Oncology	
Title: Assistant Professor	
Phone/Pager: 202-6873758	**Fax:** 202-7843034
Email address: cal9@georgetown.edu	
Mailing Address: Georgetown University, S-153, 3800 Reservoir Rd. NW, Washington DC 20057	
Study Coordinator (member of faculty or administrative official) Alexandra Schopf	

Title of Project	**Purpose of Project (one or two sentences)**
Molecular Epidemiology of Prostate Cancer	This study has two goals: 1. To establish a prostate cancer data and tissue repository; and 2. To utilize the repository to test whether prostate cancer is related to interindividual variability in the response to genotoxic stress.

Consultants, if any	**Department or Institution**
Asim Amin, M.D.	Medicine and Oncology, Georgetown University
Anatoly Dritschilo, M.D.	Radiation Medicine, Georgetown University
John Lynch, M.D.	Urology, Georgetown University
Peter Shields, M.D.	Oncology, Georgetown University
Bhaskar Kalakouri, M.D.	Pathology, Georgetown University
Mohan Verghese, M.D.	Radiation Oncology, Washington Hospital Center
Michael Porrazzo, M.D.	Urologic Oncology, Washington Hospital Center
Pamela Randolph, M.D.	Medical Oncology, Washington Hospital Center

Estimated duration of total project	3 years
Estimated total number of subjects (including control subjects)	600
Age range of subjects	>18
Sex of subjects	Male

Where will study be conducted?	GUMC
Source of subjects	Georgetown University Hospital and Washington Hospital Center

Grant Support for Project (if any)	**Commercial Support (if any) for Project**
Funded in part by the Department of Defense. Additional funding will be provided by the Lombardi Cancer Center and the protocol will be conducted by the GCRC laboratory. Once pilot data is obtained, additional grant funding will be sought.	

Investigational New Drug (IND)	**Investigational Device Exemptions (IDE)**
❑ None ❑ IND: FDA No. _____ ❑ Drug Name: _____ ❑ Drug Sponsor: _____	❑ None ❑ IDE: FDA No. _____ ❑ Device Name: _____ ❑ Device Sponsor: _____ ❑ Significant (SR) ❑ Non-Significant Risk (NSR)

Section Two: Additional MedStar Research Institute-Georgetown University Regulatory Information

1. Does this project involve the use of biohazardous materials, recombinant DNA and/or gene therapy?
   - ❑ Yes. If so, Institutional Biosafety Committee (IBC) approval must be obtained. Contact 202-687-4712 for assistance.
   - √ No.

2. Has the Institutional Biosafety Committee approved the protocol?

√ NA

❑ **Approved**	**Date Approved:**
❑ **Application Pending**	**Date Submitted:**

3. Does this project include the use of radioisotopes and/or radiation-producing devices regardless of whether the use is incidental to the project?
   - ❑ Yes. If so, all protocols must be submitted to the GUH RSC along with a completed RSC-4 or RSC-5 form. The forms require information on the use of radioisotopes and radiation-producing devices and must include dose calculations. Call 202-687-4712 to obtain forms or if additional information is required.
   - ❑ No.

4. Has the Radiation Safety Committee approved the protocol?

√ NA

❑ **Approved**	**Date Approved:**
❑ **Application Pending**	**Date Submitted:**

5. Does this project involve the use of fetal tissue?
   - ❑ Yes
   - √ No

6. Do any investigators or co-investigators have a conflict of interest as defined in the Georgetown University Faculty handbook or MedStar Health Institute policy?
   - ❑ Yes. If yes, please explain.
   - √ No

7. A copy of each investigator's current Conflicts of Interest Disclosure Form must be attached to this application.

If this project involves a FDA regulated drug or device, you must file a FDA form 3455.

Section Three: Information for Protocol Review
Please answer each specific question and use additional sheets as needed. A response of "See attached protocol or grant application" is not sufficient.

6. Provide a brief historical background of the project with reference to the investigator's personal experience and to pertinent medical literature. Use additional sheets as needed.

Despite the fact that prostate cancer is the most common tumor among US males, relatively little is known about the causative mechanisms. The known risk factors include age, ethnicity or race, high-fat diet and family history of prostate cancer, but these factors are not sufficient for identification of men with increased susceptibility. Establishing new biomarkers of cancer risk would greatly benefit the field of prostate cancer prevention and surveillance.

Mutagen sensitivity and comet assay are established biomarkers of risk (1). The mutagen sensitivity assay measures response to a genotoxic insult (e.g. bleomycin exposure) in short-term cultured human lymphocytes in terms of the number of chromatid breaks; comet assay measures DNA unwinding under alkaline conditions. Subjects with a high number of chromatid breaks in mutagen sensitivity assay or high DNA unwinding in comet assay have higher cancer risk. For example, comparison of cancer risk in the highest/lowest quartile of mutagen sensitivity in a study of 150 head and neck cancer cases and 150 controls matched on age and race showed an odds ratio of 4.5 with p=0.04 (2). Surprisingly, these phenotypic assays were not yet examined in prostate cancer. Even though the exact mechanism underlying the phenotypes is unknown, variability in DNA-repair capacity is consistent with the available experimental results (3). Moreover, it was shown in twin studies that mutagen sensitivity is heritable in non-cancer subjects. The correlation coefficient was 0.79 (95% confidence interval = 0.65-0.88) in monozygotic twins while for dizygotic twins the coefficient was 0.42 (95% confidence interval = 0.00-0.71) (4). Mutagen sensitivity and comet assay phenotypes therefore reflect multiple genetic traits related to DNA repair capacity, which predispose an individual to cancer risk.

Apoptosis is a molecular pathway eliminating, besides other functions, cells unable to cope efficiently with genotoxic stress. Deficient apoptosis is a likely candidate for a cancer-prone phenotype. Apoptosis was implicated in regulation of response to radiation therapy in prostate cancer (5), malignancy of prostatic tumor (6), and recurrence of prostate carcinoma following surgery (7). For example, in 54 prostate cancer patients treated with radiotherapy the response was negative in 84% cases with positive bcl-2 immunohistochemistry and bcl-2 was an independent prognostic variable for treatment with odds ratio of 7.3 (5). Apoptotic index was associated with disease recurrence in a study of 47 men following radical prostatectomy (7). But apoptosis was not yet examined as a phenotypic predictor of prostate cancer risk. Since the apoptotic phenotype is a composite measure of a number of converging mechanistic pathways, it is advantageous to the measurement of each individual genotype in the pathway.

Lipid peroxidation was suggested as a mechanism underlying the association of dietary fat and prostate cancer risk. Lipid peroxidation leads to oxidative genotoxic stress, that can overwhelm DNA repair and/or apoptotic mechanisms and potentially lead to cancer. We propose to quantify malondialdehyde deoxyguanosine adducts (dGMDA) in peripheral blood lymphocytes and prostate tumors. HPLC methods will be used for all assays.

DNA repair consists of two major categories, excision repair (base excision repair and nucleotide excision repair) and recombination repair (homologous and non-homologous) (8). Numerous polymorphisms in the DNA repair genes have been identified (9) and are likely to contribute to cancer risk through decreased efficiency of response to genotoxic stress. But two functional polymorphisms in DNA repair genes, *OGG1* and *XRCC1*, are particularly relevant to this study. Both genes are involved in the repair of 8-hydroxy-guanine (8-OHdG) and other oxidative lesions (10); and our study examines mainly how variability in the response to oxidative DNA damage modifies risk for prostate cancer

(bleomycin is a radiomimetic which induces oxidative DNA damage and mutagen sensitivity is mainly a model of this pathway). OGG1 is a DNA glycosylase/AP lyase involved in base excision repair of 8-OHdG and XRCC1 is a DNA ligase III terminating the base excision repair cascade (10). The OGG1 Ser(321)Cys polymorphism codes for a protein with a lower 8-OHdG repair capacity and leads to several splicing variants of unknown functional significance (11). This variant occurs at a frequency of 0.4 in Japanese and was associated with an increased risk of lung cancer in a study of 241 cases and 197 controls with an OR=3.01 (95% CI 1.33-6.83) (12). This variant was found in a Caucasian population at a frequency of 0.22 and was not associated with lung cancer in this study (13). Examination of this polymorphism in prostate cancer is therefore highly relevant. The XRCC1 Arg(399)Gln polymorphism was associated with increased sensitivity of human lymphocytes to DNA damage (14), increased risk of squamous cell carcinoma of the head and neck (15), increased risk of early onset colorectal carcinoma (16), and increased risk of adenocarcinoma of the lung (17). The polymorphism occurs in 37% of Caucasians and 17% of African-Americans (19). An examination of the XRCC1 'at risk' polymorphism as a risk factor for prostate cancer was not reported.

The study of mutations in human tumors and experimental models is elucidating important carcinogenic mechanisms (20). The study of mutations in the p53 tumor suppressor gene is uniquely suited for the study of cancer etiology, because p53 is involved in many cellular processes (including maintenance of genomic stability, programmed cell death, and DNA repair) and in tumors often accumulates point mutations amenable to further analysis (21). Specific mutations in p53 can reflect carcinogenic insults that precede cancer. It was shown that reactive oxygen species are a major source of G:C -> A:T transitions at non-CpG sites. For example, in radiation-induced lung cancer, G:C -> A:T transitions at non-CpG sites dominate the p53 mutational spectra, which differs markedly from mutational spectra associated with tobacco (22,23). Oxidatice damage is expected to be a major source of DNA damage in prostate cancer. Mutagen sensitivity and comet assay are a model of oxidative DNA damage (bleomycin is a radiomimetic which induces oxidative DNA damage), and *OGG1* and *XRCC1* participate in the repair of oxidatively damaged DNA. We therefore predict that G:C -> A:T transitions at non-CpG sites will correlate with mutagen sensitivity/comet assay phenotypes and at risk variants of *OGG1* and *XRCC1*. This study would provide for the first time an evidence for such an association. The p53 gene is also an attractive target because it is mutated in up to 35% of early prostate cancers (24).

Significance: We are proposing a molecular epidemiology study to test variation in the response to genotoxic stress and in DNA repair as a biomarker of prostate cancer risk. This study measures mutagen sensitivity, comet assay, apoptosis, and polymorphism in *OGG1* and *XRCC1* as biomarkers of prostate cancer risk; the study also correlates mutations in p53 tumor supressor gene with mutagen sensitivity. The proposal is innovative because neither of the proposed biomarkers was to our knowledge examined in connection with prostate cancer risk. If mutagen sensitivity, apoptosis, or DNA repair-variants correlate with prostate cancer risk, they could serve as readily obtainable biomarkers to identify men with increased risk of prostate cancer. The phenotypic biomarkers could be used to better identify the currently poorly understood genotoxic insults leading to cancer risk (improved risk models in case-control studies). Elucidating mechanisms of the early stages of prostate carcinogenesis would have an immediate impact for prevention and surveillance. Better prevention strategies (including chemoprevention) could be designed and tested based on the identified targets. And new hypotheses focusing on the genetic and environmental factors associated with prostate cancer risk could be formulated and evaluated.

Dr. Radoslav Goldman, Principal Investigator*:* Dr. Goldman is Assistant Professor of Oncology and a member of the Cancer Genetics and Epidemiology Program at LCC. He is an analytical toxicologist with specialization in biomarker studies of cancer risk. Dr. Goldman will be responsible for the design and execution of the proposed study, data analysis, and result interpretation. He will work in close collaboration with Dr. Loffredo and Dr. Shields on the establishment of the prostate biomarker resource.

Dr. Christopher Loffredo, Co-Investigator*:* Dr. Loffredo is Assistant Professor of Oncology and a member of the Cancer Genetics and Epidemiology Program at LCC. He is responsible for the

epidemiological field activities of the Biomarker Core Resource. Dr. Loffredo will assist with the coordination of the collection and transfer of specimen, repository, and statistical analyses.
Dr. Asim Amin, Consultant: Dr. Amin is Assistant Professor of Medicine and Oncology. He will refer patients from this department to the study coordinator.
Dr. Anatoly Dritschilo, Consultant: Dr. Dritschilo is Professor and Chairman of the Department of Radiation Oncology and will refer patients from this department to the study coordinator.
Dr. John Lynch, Consultant: Dr. Lynch is Professor of Surgery and Chairman of the Department of Urology. He will refer patients from this department to the study coordinator.
Dr. Peter Shields, Consultant: Dr. Shields is Professor of Oncology and Medicine, Director of Cancer Genetics and Epidemiology Division, and Associate Director for Population Sciences. Dr. Shields will assist in the design and oversight of the study.
Dr. Bhaskar Kalakouri , Consultant: Dr. Singh is Assistant Professor of Pathology and will oversee the collection and processing of prostate tissue for this study.
Dr. David Perry, Consultant: Dr. Perry is Medical Director of Clinical Research, Washington Hospital Center, and will refer patients to the study and help us coordinate recruitment effort at this hospital.
Dr. Mohan Verghese, Consultant: Dr. Verghese is from the Department of Radiation Oncology, Washington Hospital Center, and will refer patients from this department to the study coordinator.
Dr. Michael Porrazzo, Consultant: Dr. Porrazzo is from the Department of Urologic Oncology, Washington Hospital Center, and will refer patients from this department to the study coordinator.
Dr. Pamela Randolph, Consultant: Dr. Randolph is from the Department of Medical Oncology, Washington Hospital Center, and will refer patients from this department to the study coordinator.

7. The plan of study. State the hypothesis or research question you intend to answer. Describe the research design and procedures (including standard procedures) to be used in the research. Specifically identify any experimental procedures. Provide statistical justification for the number of subjects to be studied and the degree of change expected. Describe any special equipment or unusual procedures to be used for this research project. Use additional sheets as needed.

Research Question: This study has two goals: 1. To establish a prostate cancer data and tissue repository; and 2. To utilize the repository to test our hypothesis that prostate cancer is related to interindividual variability in the response to genotoxic stress. We propose to examine 1. Mutagen sensitivity, comet assay, and apoptotic response to bleomycin in peripheral blood lymphocytes; 2.; dGMDA adduct in lymphocytes and prostate tissue and 3. Genetic variants of the DNA repair genes *OGG1* and *XRCC1* as biomarkers of prostate cancer risk. In selected cases, we will examine the association of p53 mutational spectrum with mutagen sensitivity and genetic polymorphisms in *XRCC1* and *OGG1*.

Specific Aims: This study can address several areas of prostate cancer by developing the infrastructure to allow us to identify new biomarkers of prostate cancer risk, and improve our ability to optimize prevention and treatment strategies for prostate cancer. We plan to develop an ongoing recruitment of prostate cancer cases so that we can study prostate tumor tissue, blood and other specimen in order to understand the genotypic and phenotypic expression (e.g., mutagen sensitivity) of possible prostate cancer risk markers and to establish genotype-phenotype relationships. By linking an epidemiological profile to the tissue tumor markers, we will be able to elucidate gene-environment interactions by performing a case-control analysis and searching for etiological clues in the tumor tissue (e.g. p53 mutational spectra). The genetic risk markers under study will be limited to low penetrance genes that modulate the risk of prostate cancer and carry a risk in the context of prostate cancer of about 2-fold.

The specific aims and hypotheses of this project are to:

1. Recruit prostate cancer cases and controls to provide an epidemiological profile, blood, urine, nail clipping, and tumor tissue (when available). This will establish a data and tissue repository.

2. Utilize the repository to study low penetrance genes, investigate gene-environment interactions and establish genotype-phenotype relationships involving DNA damage, DNA repair and response to DNA damage, in order to identify or validate the use of intermediate biomarkers of cancer risk.

H_{2a} High mutagen sensitivity/comet assay increase the risk of prostate cancer.
H_{2b} Low apoptotic response increases the prostate cancer risk.
H_{2c} High dGMDA adducts increase prostate cancer risk.
H_{2d} At risk variants of XRCC1 and OGG1 increase prostate cancer risk.

3. To identify the relationship of biomarkers measured in surrogate tissues such as blood, buccal swabs and urine to pathological markers in prostate tumor. Investigate gene-environment interactions and establish genotype-phenotype relationships involving DNA damage, and response to DNA damage, in order to identify or validate the use of intermediate biomarkers of cancer risk.

H_{3a} Comet assay/dGMDA in lymphocytes correlate with these markers in prostate tissue.
H_{3b} Genetic polymorphism of DNA repair-genes is associated with p53 mutations.
H_{3c} Mutagen sensitivity is associated with p53 mutations.

Methods: Cases will be enrolled from the Departments of Medicine and Oncology, Radiation Medicine, and Urology at the Georgetown University Medical Center and Washington Hospital Center.

Approximately 200 newly diagnosed patients with prostate cancer are treated currently each year at each clinic, which is more then enough for our goal to enroll 300 patients in three years. All participants will be requested to complete an informed consent and undergo a forty five minute interview, phlebotomy, buccal cell collection and provide a nail clipping and urine sample. Also unneeded pathological tissue from patients (tumor and adjacent normal tissue) will be collected if available. A repository will be established for future studies as new hypotheses are generated.

The weekly schedule for the clinic is available to the phlebotomist/interviewer so that he/she can determine the times when eligible patients are in the clinic. Most such patients are seen at the clinic once or twice prior to their surgery so there is ample opportunity to enroll them prior to any treatment. Dr. Amin and the other consultants will inform the patients about the study and those who are potentially interested will meet the phlebotomist/interviewer. If a subject refuses to participate, then he is given the "Questions for Decliners" form and no further contact is made. The study coordinator explains the study, determines eligibility, obtains informed consent, and if appropriate administers a questionnaire, withdraws 45 cc of blood, collects buccal cells, obtains nail clipping and a urine sample in collaboration with the GCRC laboratory. As the patients await their examination in the clinic, they are accompanied by the phlebotomist/interviewer who helps them with orientation in the building etc. This gives also opportunity to answer the preliminary questions and to set a time for the full questionnaire/sample collection. This method worked well in our previous studies.

Controls are obtained from visitors accompanying other patients to the hospital. The interviewer identifies potential candidates, investigates their willingness to participate, and screens for eligibility using a script (Script 2-Control Recruitment in Clinic Area) and the eligibility screening form. The subjects usually accompany a person to the hospital on a regular basis. These controls are easily contacted and typically motivated to participate. The interviewer creates a list of willing, eligible controls and recruits from the list to the study when a match is identified. The controls are unbiased with respect to geography and socioeconomic status because they come to the hospital from the same geographic referral area as the cancer cases. In addition, controls can be obtained from neighbors and friends of the patients. Each patient can nominate up to 5 people living in the same geographical area and of the same race and age (within 5 years). The patients are asked to verify with the nominees about their agreement to be contacted by the phlebotomist/interviewer. A random drawing from the list of candidates will be performed and a candidate will be contacted. Up to three phone calls will be placed. If the subject does not return the phone calls, then it is assumed that he is uninterested in participating. In the event that a subject cannot be reached by phone, he will be contacted by mail. In case of refusal, next candidate is then randomly selected from the list of nominees. An attempt is made to collect information on age, race, smoking and drinking history of those who refuse to participate to determine whether they differ from participants demographically or by exposures. If a matching control cannot be found among the nominees, a match is identified from the pool of all eligible controls in the study. The phlebotomist/interviewer works from a list of the cases that have been enrolled up to that time, so that he/she can identify appropriate matches. Eligibility of interested controls to participate is determined over the phone by the phlebotomist/interviewer according to the telephone script. The interested candidates are invited to the Georgetown Hospital to finish a full questionnaire, donate a 45cc blood sample, a sample of buccal cells, and a sample of urine. PSA will be tested by the GCRC for all controls to exclude misclassification. Controls with PSA > 2.5ng/ml will be referred to a clinician for a follow-up testing. In this way, we obtain controls individually matched on race and age (within 5 years). Informed consent is obtained at the time of interview.

Additionally, all men undergoing a prostate biopsy at GUMC will be given a "consent to participate in research" form. Of those that consent to participate in research, the patients whose biopsy is positive will be recruited into the Case group, while the patients whose biopsies are negative will be recruited into the Control group. This control group of men with confirmed negative biopsies will constitute a group of men with Benign Prostatic Disease, and will be a separate control group from those who have no diagnosis of

prostatic disease and have never received a biopsy.
It should be noted that representatives of the U.S. Army Medical Research and Materiel Command are eligible to review research records as part of their responsibility to protect human subjects in research. Also, if any changes to the protocol or consent form are made, they are to be reviewed and approved by the Human Subjects Research Review Board prior to implementation.

Reporting of Serious and Unexpected Adverse Events:
Unanticipated problems involving risk to subjects or others, serious adverse events related to participation in the study, and all study-related subject deaths will be promptly reported by phone (301-619-2165), by email (hsrrb@det.amedd.army.mil), or by facsimile (301-619-7803) to the Army Surgeon General's Human Subjects Research Review Board (HSRRB). A complete written report will follow the initial telephone call. In addition to the methods above, the complete report can be sent to the U.S. Army Medical Research and Materiel Command, ATTN:MCMR-ZB-QH, 504 Scott Street, Fort Detrick, Maryland 21702-5012."
Procedures: Subjects are identified by review of appointment logs and discussion with doctors. Subjects are contacted during their visit to the clinic (patients), in the clinic waiting areas (controls), or by phone (controls nominated by the patient). The phlebotomist/interviewer assists the patient during his visit to the hospital, determines eligibility, explains the study and obtains informed consent, administers the questionnaire and collects 45cc of blood, buccal cells, nail clipping and a sample of urine together with the GCRC laboratory. The interviewers are trained through the GCRC in how to administer and properly complete the questionnaire. Dietary exposures (high fat etc.) will be assessed using the well-validated questionnaire developed by Dr. Gladys Block, NCI, NIH. Phlebotomy is performed by trained phlebotomists. There will be a single blood draw, using these tubes in the following order: two 7 ml green top tubes, two 7 ml plain red top tubes, one 10 ml yellow top tubes, and one 7 ml purple top tube. Only a portion of the collected samples is used for the currently planned specific aims. The remainder of the samples is aliquotted and frozen at -70^{o}C for future studies. There will be blood for multiple aliquots of buffy coat, mononuclear cells, PMNs, serum, plasma, red blood cells and clots. This strategy will allow us to test new hypotheses and assess new genetic predispositions as they are deemed worthy of study. If the subject is going to surgery, residual normal and tumor prostate tissue is placed into aliquots and snap frozen. Two samples of the normal and tumor tissues is saved, one without preservative and one with RNA later for preserving RNA. Tumor tissue is also fixed in formalin and ethanol. When available from surgery, normal cells are collected to establish primary cell cultures. If a subject is not going to surgery, but the subject had surgery at the University, then tumor blocks are requested from the LCC histopathology core. Medical records are reviewed to obtain pathological and clinical data. If a subject chooses to withdraw from the study, the link between his identity and the research study will be destroyed. Also, his biological samples will be discarded. However, the results of any finished analysis and or published result will be kept to preserve the integrity of the study.

Laboratory Methods: All the methods follow an established protocol. The mutagen sensitivity, comet assay, and apoptosis are carried out on short-term (3 day) cultured human lymphocytes exposed to bleomycin (2). The samples of isolated DNA for dGMDA quantification are sent to outside collaborators for analysis. These samples will contain only the identifier code so that there is no possibility to disclose personal information. The dGMDA is quantified by gas chromatography/negative chemical ionization mass spectrometry (25). Genetic polymorphisms are analyzed by PCR-RFLP as described (12)(19). Mutational spectra of p53 are analyzed in isolated DNA by the affymetrix chip in the laboratory of Dr. Shields (26).
Statistical Power: The present proposal intends to study 300 prostate cancer cases and 300 matched controls. The matched-pairs design increases statistical power to detect a meaningful relative risk since matched-pairs data would gain relative efficiency in estimation. Suppose the hypothesis of interest is that

having a certain biomarker (e.g. mutagen sensitivity) increases the probability of developing prostate cancer, with the null hypothesis being that such probability is the same with or without the biomarker. Let *p* be the population frequency of having such biomarker, and let *r* be the relative risk defined as the ratio of the frequency of prostate cancer with the biomarker to the frequency of prostate cancer without the biomarker. Then for *r*=2.5, the statistical power with 5% level of significance (two-sided) will be 84%, 89%, and 93%, respectively, if *p*=20%, 25%, and 30%, accordingly. In our case, for example, the frequency of mutagen sensitive subjects in the population was estimated as 20% (6) and the *XRCC1* 'at risk' allele as 25% in the general population (19). The statistical power would be relatively lower when the comparison is controlled by other factors such as race. It should be noted that tests of effect modification or associations are exploratory, and the study was not designed to have optimal power for those analyses. All the analyses will be performed using the Statistical Analysis System (SAS) and S-plus statistical software packages.

References:

1. Berwick, M. and Vineis, P. Markers of DNA Repair and Susceptibility to Cancer in Humans: an Epidemiologic Review. J Natl Cancer Inst. 6-7-2000;92(11):874-97.
2. Wu, X., Gu, J., Patt, Y., Hassan, M., Spitz, M. R., Beasley, R. P., and Hwang, L. Y. Mutagen Sensitivity As a Susceptibility Marker for Hepatocellular Carcinoma. Cancer Epi.Biom.Prev. 1998;7(7):567-70.
3. Wei, Q., Spitz, M. R., Gu, J., Cheng, L., Xu, X., Strom, S. S., Kripke, M. L., and Hsu, T. C. DNA Repair Capacity Correlates With Mutagen Sensitivity in Lymphoblastoid Cell Lines. Cancer Epidemiol Biomarkers Prev. 1996;5(3):199-204.
4. Cloos, J., Nieuwenhuis, E. J., Boomsma, D. I., Kuik, D. J., van der Sterre, M. L., Arwert, F., Snow, G. B., and Braakhuis, B. J. Inherited Susceptibility to Bleomycin-Induced Chromatid Breaks in Cultured Peripheral Blood Lymphocytes . J.Natl.Cancer Inst. 7-7-1999;91(13):1125-30.
5. Scherr, D. S., Vaughan, E. D., Wei, J., Chung, M., Felsen, D., Allbright, R., and Knudsen, B. S. BCL-2 and P53 Expression in Clinically Localized Prostate Cancer Predicts Response to External Beam Radiotherapy. J Urol. 1999;162(1):12-6.
6. Lipponen, P. and Vesalainen, S. Expression of the Apoptosis Suppressing Protein Bcl-2 in Prostatic Adenocarcinoma Is Related to Tumor Malignancy. Prostate 6-15-1997;32(1):9-15.
7. Stapleton, A. M., Zbell, P., Kattan, M. W., Yang, G., Wheeler, T. M., Scardino, P. T., and Thompson, T. C. Assessment of the Biologic Markers P53, Ki-67, and Apoptotic Index As Predictive Indicators of Prostate Carcinoma Recurrence After Surgery. Cancer 1-1-1998;82(1):168-75.
8. Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. Human DNA Repair Genes. Science 2001;291:1284-1289.
9. Shen, M. R., Jones, I. M., and Mohrenweiser, H. Nonconservative Amino Acid Substitutions Exist at Polymorphic Frequency in DNA Repair Genes in Healthy Humans. Can. Res. 2-15-1998;58:604-8.
10. Boiteux, S.,and Radicella, J. P. The Human OGG1 Gene: Structure, Function, and its Implication in the Process of Carcinogenesis. Arch. Biochem. Biophys. 2000;377(1):1-8.
11. Kohno, T., Shinmura, K., Tosaka, M., Tani, M., Kim, S. R., Sugimura, H., Nohmi, T., Kasai, H., and Yokota, J. Genetic Polymorphisms and Alternative Splicing of the hOGG1 Gene that is Involved in Repair of 8-hydroxyguanine in Damaged DNA. Oncogene. 1998;16(25):3219-3225.
12. Sugimura, H., Kohno, T., Wakai, K., Nagura, K., Genka, K., Igarashi, H., Morris, B. J., and Yokota J. hOGG1 Ser326Cys Polymorphism and Lung Cancer Susceptibility. Cancer Epidemiol Biomarkers Prev. 1999;8(8):669-674.
13. Wikman, H., Risch, A., Klimek, F., Schmezer, P., Spiegelhalder, B., Dienemann, H., Kayser, K., Schulz, V., Drings, P., and Bartsch, H. hOGG1 Polymorphism and Loss of Heterozygosity (LOH): Significance for Lung Cancer Susceptibility in a Caucasian Population. Int J Cancer. 2000;88(6):932-937.
14. Duell, E. J., Wiencke, J. K., Cheng, T. J., Varkonyi, A., Zuo, Z. F., Ashok, T. D., Mark, E. J., Wain, J. C., Christiani, D. C., and Kelsey, K. T. Polymorphisms in the DNA Repair Genes XRCC1 and ERCC2 and Biomarkers of DNA Damage in Human Blood Mononuclear Cells [Published Erratum Appears in Carcinogenesis 2000 Jul;21(7):1457]. Carcinogenesis 2000;21(5):965-71.
15. Sturgis, E. M., Castillo, E. J., Li, L., Zheng, R., Eicher, S. A., Clayman, G. L., Strom, S. S., Spitz, M. R.,

and Wei, Q. Polymorphisms of DNA Repair Gene XRCC1 in Squamous Cell Carcinoma of the Head and Neck. Carcinogenesis 1999;20(11):2125-9.

16. Abdel-Rahman, S. Z., Soliman, A. S., Bondy, M. L., Omar, S., El Badawy, S. A., Khaled, H. M., Seifeldin, I. A., and Levin, B. Inheritance of the 194Trp and the 399Gln Variants of the DNA Repair Gene XRCC1 Are Associated With Colorectal Carcinoma in Egypt. Cancer Lett. 2000;159(1):79-86.
17. Divine, K. K., Gilliland, F. D., Crowell, R. E., Stidley, C. A., Bocklage, T. J., Cook, D. L., and Belinsky, S. A. The XRCC1 399 Glutamine Allele Is a Risk Factor for Adenocarcinoma of the Lung. Mutat.Res 1-5-2001;461(4):273-8.
18. Block, G., Patterson, B., and Subar, A. Fruit, Vegetables, and Cancer Prevention: a Review of the Epidemiological Evidence. Nutr.Cancer 1992;18:1-29.
19. Lunn, R. M., Langlois, R. G., Hsieh, L. L., Thompson, C. L., and Bell, D. A. XRCC1 Polymorphisms: Effects on Aflatoxin B1-DNA Adducts. Cancer Res. 6-1-1999;59(11):2557-61.
20. Dogliotti, E., Hainaut, P., Hernandez, T., D'Errico, M., and DeMarini, D. M. Mutation spectra resulting from carcinogenic exposure: from models to cancer genes. Rec. Res. Can. Res., *154*: 97-124, 1998.
21. Levine, A. J. p53, the cellular gatekeeper for growth and division. Cell, *88*: 323-331, 1997.
22. DeBenedetti VMG, Travis LB, et al. p53 mutations in lung cancer following radiation therapy for hodgkin's disease. Cancer Epidemiol Biomark Prev 1996; 5:93-98.
23. Vahakangas KH, Samet JM, et al. Mutations of p53 and ras genes in radon-associated lung cancer from uranium miners. Lancet, 1992; 339:576-580.
24. Chi SG, deVereWhite RW, et al. p53 in prostate cancer: frequent expressed transition mutations. JNCI 1994; 86:926-933.
25. Otteneder M, Plastaras JP, and Marnett LJ. Reaction of malondialdehyde DNA adducts with hydrazines-a facile assay for quantification of malondialdehyde in DNA. Chem Res Tox. 2002 Mar; 15:312-318
26. Wolf P, Hu YC, Doffek K, Sidransky D, Ahrendt SA. O(6)-Methylguanine-DNA methyltransferase promoter hypermethylation shifts the p53 mutational spectrum in non-small cell lung cancer. Cancer Res. 2001 Nov 15;61(22):8113-7.

8. Indicate what you consider to be the risks to subjects and indicate the precautions to be taken to minimize or eliminate these risks. Justify the need for a placebo control group if one is included in this study. Where appropriate, describe the data monitoring procedures that will be employed to ensure the safety of subjects. Use additional sheets as needed.

There are minimal risks for this study. The only invasive procedure is phlebotomy. This may cause a bruise on the arm from the needle stick and possibly an infection. These risks are minimized through proper techniques for phlebotomy and the trained staff is experienced in reducing discomfort to patients. The actual surgery or clinical practices related to the prostate cancer will not be altered for this study.

Section Four: Selection of Subjects and the Informed Consent Process

9. Indicate whether this project involves any of the following subject populations?
   - ❑ Children (Children are defined by local law as anyone under age 18.)
   - ❑ Prisoners
   - ❑ Pregnant women
   - ❑ Cognitively impaired or mentally disabled subjects
   - ❑ Economically or educationally disadvantaged subjects

If you indicated any of the above, in the space below, please describe what additional safeguards will be in place to protect these populations from coercion or undue influence to participate. (Use additional sheets as needed.)

10. Describe how subjects will be recruited and how informed consent will be sought from subjects or from the subjects' legally authorized representative. If children are subjects, discuss whether their assent will be sought and how the permission of their parents will be obtained. Use additional sheets as needed.

This is a study of prostate cancer risk factors that enrolls newly diagnosed, incident prostate cancer cases from the Departments of Medicine and Oncology, Radiation Medicine, and Urology at the Georgetown University Medical Center. The eligible patients donate their time for a questionnaire; blood and urine samples; buccal swabs; nail clipping; and unneeded normal and tumor prostate tissue. Subjects are eligible and will be enrolled even if they are not having a surgery or biopsy and if no tissues are available. Subjects older than 18 years of age at all stages of presentation are included. No subject is excluded based on minority status. Subjects with psychiatric disorder or any other reason that precludes understanding the informed consent are excluded for ethical reasons. The phlebotomist/interviewer conducts a brief initial 15 minute interview in order to explain the study, determine eligibility, and explain the informed consent. If a subject refuses to participate, then no further contact is made. If appropriate, the phlebotomist/interviewer administers a structured forty five minute interview that establishes demographic characteristics, family history of cancer, dietary habits, tobacco and alcohol use, occupational exposures, and history of vasectomy. This interview can be done at any time up to two months after initiation. The phlebotomist/interviewer will also withdraw 45 cc of blood, collect buccal cells, obtain nail clipping and a urine sample in collaboration with the GCRC laboratory at Georgetown University.

Controls are obtained from visitors accompanying other patients to the hospital. The interviewer identifies potential candidates, investigates their willingness to participate, and screens for eligibility using a one-page form. The interviewer creates a list of willing, eligible controls and recruits from the list to the study when a match is identified. In addition, controls can be obtained from neighbors and friends of the patients. Each patient can nominate up to 5 people living in the same geographical area and of the same race and age (within 5 years). The patients are asked to verify with the nominees about their agreement to be contacted by the phlebotomist/interviewer. The controls are randomly selected from the list of candidates and contacted by the interviewer. Up to three phone calls are placed. If the subject does not return the phone calls, then it is assumed that he/she is uninterested in participating. In case of refusal, next candidate is randomly selected from the list of nominees. An attempt is made to collect information on age, race, smoking and drinking history of those who refuse to participate to determine whether they differ from participants demographically or by exposures. A subsequent meeting with the matching

control is scheduled. During this meeting, the interviewer explains the study in detail and obtains informed consent. A full length questionnaire as well as blood , buccal, urine, and nail-clipping samples are obtained. The samples or questionnaire can be obtained also at a later visit up to two month following the initial contact if this is more convenient for the participant.

11. Will subjects receive any compensation for participation in cash or in kind?
    - √ Yes. If so, please describe amount or kind of compensation in the space below.
    - ❑ No.

Patients will not be compensated. Controls will receive free PSA test if needed and $25 for parking if study funds permit.

Section Five: Privacy and Confidentiality of Data and Records

12. Will identifiable, private, or sensitive information be obtained about target the subjects or other living individuals? Whether or not such information is obtained, describe the provisions to protect the privacy of subjects and to maintain the confidentiality of data. Use additional sheets as needed.

There are minimal risks of disclosure of sensitive information in this study, but there is always the risk that genetic or other risk factor data might be obtained by the subject or a third party. However, it is important to realize that the genes studied herein are low penetrant. We study only common genetic polymorphisms in DNA repair genes and somatic mutations in p53; we do not study familial germ line mutations. This risk of disclosure will be minimized by the confidentiality and protection of privacy procedures described below.

Protection of privacy of participants in genetic studies is of the utmost importance. Study subject's confidentiality is maintained at all times. Subjects are assigned unique study numbers. These unique study numbers are linked to the subject's identifier information in a database and on the hard copy of the Identifier Sheet. This information is secured by Dr. Goldman in his office separate from the laboratory. The database requires at least two levels of security (i.e. passwords), which allows only authorized individuals to access the information. The Identifier Sheets are physically separated from the questionnaire and stored in a locked cabinet. The questionnaire retains only the unique study number. Biological samples are labeled with the unique study number and no other identifier information. No identifier information that can be linked to study results or other data will leave Dr. Goldman's premises.

Identifier information for non-participants (refusers and ineligibles) is recorded in order to avoid recontact. This information is stored in a database with at least two levels of security (i.e. passwords), which allows only authorized individuals to access the information. A log will automatically note who accesses the information and what was accessed. Unique study number for non-participants is also assigned; this is used for tracking reasons. Two databeses are maintained. The first includes the Contact Database and includes identifier information. It will record if subjects refused, were ineligible, or are participants. If participants, it will record when the interview occurred or will occur, the outcome, and track sample handling. For refusers and ineligibles, it will record that their data was entered into the Refusal and Ineligible database. The Refusal and Ineligible database will record data and why the subject was ineligible. This database does not contain identifier information.

I certify that the information furnished concerning the procedures to be taken for the protection of human subjects is correct. I will seek and obtain prior approval for any modification in the protocol or informed consent document and will report promptly any unexpected or otherwise significant adverse effects encountered in the course of this study.

I certify that all individuals named as consultants or co-investigators have agreed to participate in this study.

_____ Printed/Typed Name of Investigator _____ Signature of Investigator	_____ Telephone number _____ Date
Department Chair: ❑ Approved ❑ Disapproved _____ Printed/Typed Name _____ Signature of Department Chair	 _____ Telephone Number _____ Date

If more than one department or administrative unit is participating in the research and/or if the facilities or support of another unit, e.g., nursing, pharmacy, or radiation therapy, are needed, then the chair or administrative official of each unit must also sign this application.

_____ Authorized Signature and Title	_____ Date
_____ Authorized Signature and Title	_____ Date
_____ Authorized Signature and Title	_____ Date
_____ Authorized Signature and Title	_____ Date
_____ Authorized Signature and Title	_____ Date

Section Six: Attachments

Please attach the following items in order for the IRB to review your research.

1. 24 copies of this IRB Application form
2. The informed consent document (24 copies)
3. Any recruitment notices or advertisements (24 copies)
4. Any research survey instruments, psychological tests, interview forms, or scripts to be used (24 copies).
5. Certificate of Completion of Education in the Protection of Human Research Subjects
6. Investigator's qualifications (CV, biosketch, or Form 1572, if available)
7. Investigator's Brochure from the sponsor, if applicable (5 Copies)
8. Research protocol and sample consent document from the sponsor or Cooperative Group, if applicable (5 copies)
9. Grant application, if applicable (2 copies)

Investigator's Brochure (where applicable)

The Investigator's Brochure must contain the following information. If it does not contain the information, then please attach a separate sheet of paper to address the item.

(a) Name of drug under study.
(b) Source of the drug.
(c) Experience with the drug in humans, including doses tested, toxicity observed, minimal toxic dose, pharmacokinetic data (absorption, elimination, metabolism, etc.).
(d) Description of toxicity in humans.
(e) Procedures for minimizing adverse reactions and dealing with those that might occur.

Mutagen Sensitivity Eligibility Criteria

Name: _____ MR# _____

1. Have you had a previous diagnosis of any cancer? ( ) Yes ( ) No
   If yes, what kind of cancer?_____

2. Have you received chemotherapy for any reason within the past 6 months?
   ( ) Yes ( ) No
   If yes, what dates? _____

3. Have you received radiation for any reason within the past 6 months?
   ( ) Yes ( ) No
   If yes, what dates? _____

4. Have you had any surgeries within the past month that required anesthesia?
   ( ) Yes ( ) No

5. Are you being treated for infection or have you taken antibiotics within the past 12 days?
   ( ) Yes ( ) No
   When will you finish your antibiotics? _____

6. Have you received a blood transfusion within the past 6 months?
   ( ) Yes ( ) No

---

7. Are you taking any steroids or immunosuppressive medications?
   ( ) Yes ( ) No
   When will you finish your medications? _____

8. Do you have a known diagnosis of HIV, hepatitis B or C?
   ( ) Yes ( ) No

9. Are you an IV drug user?
   ( ) Yes ( ) No

****

MEDICAL RECORDS RELEASE AND
GENERAL AUTHORIZATION TO USE AND DISCLOSE HEALTH INFORMATION FOR RESEARCH

I agree to allow Dr. Goldman and his staff (together called "Researchers"), as well as the study sponsor, Lombardi Cancer Center of Georgetown University, others working with the sponsor to do the research (together called "Sponsor"), and the other people or companies listed below, to use and give my personal health information that identifies me for the reason described in the Informed Consent Form used for this study and as needed to conduct the research. I also agree to allow Georgetown University Hospital, my doctors and my other health care providers, and others who generate or use my health information, to give my health information in my medical or other records to the Researchers and Sponsor for the purposes described below and in the Informed Consent Form used in this study. [IRB Project # 03013 and Project Full Title: The Molecular Epidemiology of Prostate Cancer]

1. The health information that may be used for this study includes:

☒ All my personal information made or collected during the research described in the Informed Consent Form for this study; and

☒ All my personal information in my medical records requested by the Researchers to be able to do the research described in the Informed Consent Form for this study.

OR

☐ The following information: _____

_____

2. The person(s), class(es) of persons, and/or organizations (companies) who may use, give and receive the above information include*:

☒ Every research site for this study, including the hospital, and including each site's research staff, medical staff and administrative staff;

☒ Health care providers who provide services to me in connection with this study;

☒ Laboratories and other individuals and organizations that look at my health information in connection with this study, in agreement with the study's protocol;

☐ The Sponsor and the people and companies that they use to watch over how the study is managed, run, or do the research as described above;

☐ The United States Food and Drug Administration (FDA) and other Federal or State Agencies that watch over the safety of the study and how the study is managed or run;

☒ The members and staff of the Institutional Review Board(s) or Ethics Committee(s) that approves this study;

☒ The Principal Investigator, other Investigators, Study Coordinators, and all administrative staff in charge for doing all the work for the study and other research activities;

☐ The Patient Advocate or Research Ombudsman (people who watch out for my best interest):

_____

☐ Data Safety Monitoring Boards (a group of people who examine the medical information during the study) and other government agencies or review boards who watch over the safety, success and how the research is done.

☐ Others:_____

_____

**If, during the course of the research, one or more of the companies or institutions above merges (becomes one company) or is bought by another company, this Authorization will remain valid.*

3. Once my health information has been given to one of the person(s), class(es) of persons, and/or organizations (companies) listed above, there is the possibility that federal privacy laws (laws that protect the privacy to my personal health information) may no longer protect it from being given to

another person, class of persons, and/or company. However, the Researchers and Sponsor [may agree/have agreed] to further protect my health information by using and disclosing it only for the research purposes described in the Informed Consent Form and as allowed by me in this Authorization (agreement). Also, the Researchers and Sponsor [may agree/have agreed] that no publication or presentation of the research will reveal my identity without my separate specific written permission and authorization (agreement). These limitations, if agreed to by the Researcher and Sponsor, continue even if I revoke (take back) this Authorization (agreement).

4. **Once information that could be used to identify me has been removed and my information is no longer identifiable (connected to my identity) under federal regulations,** the information that remains is no longer protected by this Authorization (agreement) and may be used and given by the Researchers and Sponsor as permitted by law to others, including for other research reasons.

5. **I understand that:**
   - I have the right to refuse to sign this Authorization (agreement). While my health care outside the study, the payment for my health care, and my health care benefits will not be affected if I do not sign this form, I will not be able to participate in the research described in this Authorization (agreement) and will not receive treatment as a study participant if I do not sign this form.
   - I may change my mind and revoke (take back) this Authorization (agreement) at any time. To take back this Authorization (agreement), I must write to: Alexandra Schopf, Lombardi Cancer Center, Lower Level Room S-157, Georgetown University, Box 571472, Washington, DC 20057-1472 . However, if I take back this Authorization (agreement), I may no longer be allowed to participate in the research. Also, even if I take back this Authorization (agreement), the information already obtained may remain a part of the research as necessary to preserve the integrity of the research study.

6. **This Authorization (agreement) does not have an expiration (ending) date.**

7. **I will be given a copy of this Authorization (agreement) after I have signed it.**

8. **I acknowledge that I have received or declined the pamphlet with the MedStar Health Notice of Privacy Practices and that this form is available for me to take with me.**

_____	_____
Signature of participant or participant's legal representative	Date
_____	_____
Printed name of participant or participant's representative	Representative's authority to sign for participant

For Internal Use Only

Signature/acknowledgement of receipt of Notice of Privacy Practices not obtained because:

- ☐ Emergency
- ☐ Patient/Patient Representative declined to sign
- ☐ Patient/Patient Representative unable to sign

_____
MRI Representative

TELEPHONE CONTACT-Prostate

- Hi my name is Alexandra Schopf and I am calling from the Lombardi Cancer Center at Georgetown University. You were referred to me by Dr…….who is conducting a research study with us here at LCC. Dr. …..suggested I contact you and ask you to participate. My colleague, Tara Lamond, may have already spoken with you regarding her study. Please understand that these are two different studies, but are complementary to each other.
- I would like to tell you a little more about this research project designed to improve our understanding of prostate cancer.
- The Study is entitled "Prostate Cancer Biomarker Resource" and is funded through the Department of Defense.
- Our objective is to provide our medical researchers with an epidemiological profile in the form of a questionnaire as well as biological samples. Thus, should you choose to participate you will first be asked to sign an informed consent form, take part in a ten minute interview and to provide a small sample of blood, urine, mouthwash and a toenail sample.
- I would just like you to know that all information is kept strictly confidential. There is no information listed on the questionnaire or biologic specimen to reveal your identity. Additionally, joining the study is completely voluntary and will have no positive or negative effect on your relationship with your doctor, treatment plans etc.
- Your participation in this study will help us test new methods for early diagnosis and treatment of prostate cancer. Such information is invaluable for both present and future patients. Does this sound like something you would be interested in participating in?
- IF NO – could I ask you why you are not interested? Also, could I ask you just a few questions? What is your occupation? Do you smoke tobacco or drink alcohol on a regular basis? (Also, find out race, DOB, and enter all information in database)

  Then-Thank you very much for your time. Best wishes for a fast recovery.

  IF YES – I just want to confirm
    1. Have you ever had cancer before?
    2. Have you had any chemotherapy or radiation in the past 6 months?
    3. Have you had any MAJOR surgeries (biopsy is not major) in the past 3 months?
- (If no to all 3 questions) OK, we can schedule an appointment to meet either before or after your next visit to GU. When is that? (or, if you would like to make a separate trip, we can pay for parking). It will take about one hour for me to explain the study, have you sign the consent forms, collect your biological samples and conduct the ten minute interview. (Agree on time and place to meet). Also, sir, please don't clip your toenails for about a week before our appointment. Thank you. See you soon.

Control recruitment protocol-approaching people in clinic waiting areas

(Interviewer carries 'matching' chart with her/him around clinic, approaches men who appear to fit the needed demographics)

- Excuse me sir (male between 18 and 90 yrs old-ask if unsure), are you a patient here?
- If cancer patient: Thank you. If patient seems curious, explain: I am working on a research study here and looking for people who are here accompanying patients.
- OTHERWISE: Hi I'm Alexandra Schopf. I'm working on a research project designed to improve our understanding of prostate cancer. Do you have a minute to hear about our study?

If NO: Ok, sorry to bother you.

If YES: Thanks. Right now, very little is known about why people get prostate cancer. We are concerned, and are currently investigating biological factors linked to prostate cancer susceptibility.

- Right now, we are looking for people who have no cancer history to participate in the study as part of a healthy comparison group for our participants who have prostate cancer. Might you be interested in participating?
- If no or 'I had (something other than skin) cancer before' : Ok, thank you for your time. Good luck with your visit today.
- If yes, continue:
- The Study is entitled "Prostate Cancer Biomarker Resource" and is funded through the Department of Defense.
- Our objective is to provide our medical researchers with an epidemiological profile in the form of a questionnaire as well as biological samples. Thus, should you choose to participate you will first be asked to sign an informed consent form, take part in a 45minute interview and to provide a small sample of blood, urine, saliva, and toenail clippings.
- I would just like you to know that all information is kept strictly confidential. There is no information listed on the questionnaire or biologic specimen to reveal your identity. Additionally, joining the study is completely voluntary and will be of no direct benefit to you, but could help us develop better methods for understanding, diagnoses and treatment of prostate cancer. Such information is invaluable for both present and future patients. Would you like to participate?
- If YES: Administer control screening form.

If person tells of previous cancer diagnosis: I am sorry I wasn't so clear earlier, we are looking to enroll people with no cancer history. Thank you very much for your time and best of luck with your visit today.

If person meets eligibility criteria: It will take about 45 minutes for me to explain the study, have you sign the consent forms, and collect your biological samples. There is also a 45 minute interview that we could do here at GU if you have time or over the phone at your convenience. If we complete the interview here the whole thing would take under two hours. There would be no follow up. It would be just a one-time commitment. Do you have time today? If not, when do you plan on returning to the clinic? (Agree on a time to meet. Otherwise hand person brochure and point out

contact info on the back. Ask them to please call when they know their schedule). See you soon.

- If person declines at any time: Can I ask why you aren't interested? (find out age, race, smoking and drinking history as well as level of education) Thank you for your time best of luck with your visit today.

Molecular Epidemiology of Prostate Cancer (Case/Control)

Principal Investigator: Radoslav Goldman, Ph.D.
Department of Oncology
Lombardi Comprehensive Cancer Center
Georgetown University Medical Center
LCC, LL (S) Level, S183
3800 Reservoir Road, NW
Washington, DC 20057
Tel: (202) 687 9868
Fax: (202) 687 1988
email: rg26@georgetown.edu

TABLE OF CONTENTS

Date of Interview ____ ____ _____ MM DD YYYY	Time of Interview _____ ☐ 1 AM _____ ☐ 2 PM

Interviewer	Interviewer Signature

Study ID/ Site ID	LCC Number

MRN	Control? Yes_____ No_____

Reviewers initials	Date reviewed ____ ____ _____ MM DD YYYY
Coders initials	Dated coded ____ ____ _____ MM DD YYYY
First Entry initials	Date entered ____ ____ _____ MM DD YYYY
Second entry initials	Date entered ____ ____ _____ MM DD YYYY

Samples Collected	ID label
Blood ☐ ___ yellow ___ red ___ green ____purple	
Mouthwash ☐	
Urine ☐	
Toenail ☐	
Tissue ☐	
Other ☐	

Your answers to the following questions are very important to us. Please answer them as truthfully as possible. Also, please remember that you do not have to answer any question that makes you feel uncomfortable.

A. IDENTIFIER SHEET

A1. What is your name? _____ / _____ / _____
First Middle Last

A2. Could your medical records be under a different name? If so, what name?

_____ / _____ / _____
First Middle Last

A3. What is your date of birth? ___ ___ / ___ ___ / ___ ___ ___ ___
MM DD YYYY

A4. What is your address?

_____
Street Apt. No.

_____ ____ ____ __ __ __ __ __ - __ __ __ __
City State Zip Code

_____
Country

A5. What is your telephone number? Home: (__ __ __) __ __ __ - __ __ __ __

Work: (__ __ __) __ __ __ - __ __ __ __

Ext. __ __ __ __

Email _____

A6. Is there someone at a different address that would be able to help us contact you in the future?

_____ _____
Name Relationship to person

_____ _____
Street Apt. No.

_____ ____ ____ __ __ __ __ __ / __ __ __ __
City State Zip code

Home Telephone Number: (__ __ __) __ __ __ - __ __ __ __

Work Telephone Number: (__ __ __) __ __ __ - __ __ __ __

Ext. __ __ __ __

Email _____

IDENTIFIER SHEET ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

B. DEMOGRAPHIC INFORMATION

Now I would like to ask you some general information about yourself.

B1. What is your marital status?

( )1 Widowed
( )2 Married or living as married
( )3 Divorced
( )4 Separated
( )5 Single, never married

B2. Which of these categories best describes you?

( )1 White
( )2 Black or African American
( )3 Asian
( )4 Native Hawaiian or Other Pacific Islander
( )5 Other Specify_____

B3. What country or continent were you born in?

( )1 United States ( )2 Africa ( )3 Europe
( )4 Caribbean/West Indies ( )5 Asia ( )6 South America
( )7 Middle East ( )8 Canada ( )9 Australia
( )10 United Kingdom ( )11 Central America ( )12 Other_____

B4. If you moved from here, at what age did you move?_____

B5. What was the highest level of education you completed (don't read choices).

( )1 Less than 8th grade ( )2 Less than high school ( )3 High school graduate
( )4 Less than 4 years of college ( )5 College (4 years completed)
( )6 Graduate/professional coursework or degree

B6. In what religion were you raised?

( )1 Protestant ( )2 Catholic ( )3 Muslim
( )4 Jewish ( )5 None ( )6 Other Specify _____

If Jewish, are you Ashkenazi? _____yes _____no

B7. What is your current level of household income per year (read choices)?

( )1 Less than $25,000
( )2 $25,001 - $50,000
( )3 $50,001 - $100,000
( )4 $100,001- $150,000
( )5 Greater that $150,000
( )8 Don't know

B8. How many people are currently supported in your household? ____ ____

DEMOGRAPHIC INFO ( )1 Very Good ( )2 Good ( )3 Fair ( )4 Poor

C. MEDICATIONS

C1. Now I have some questions about any prescription medication you may have taken.

Drugs	C1.Have you ever taken (DRUG)?	C2. In what year did you first take (DRUG)?	C3. For how long did you take (DRUG)?	C4. How often did you take (DRUG) per day or per week?
a. Propecia used to treat baldness?	YES ..... 1 → NO ..... 2 (b)	|___|___|___|___|	|___|___| MONTHS ..... 1 YEARS 2	|___|___| PER DAY ..... 1 PER WEEK 2
b. Proscar or fenasteride used to treat prostate disease?	YES ..... 1 → NO ..... 2 (c)	|___|___|___|___|	|___|___| MONTHS ..... 1 YEARS 2	|___|___| PER DAY ..... 1 PER WEEK 2
c. Luprone or Zolodex used to treat prostate disease?	YES ..... 1 → NO ..... 2 (d)	|___|___|___|___|	|___|___| MONTHS ..... 1 YEARS 2	|___|___| PER DAY ..... 1 PER WEEK 2
d. Flutamide also called Eulexin; or Nilandron; or Casodex used to treat prostate disease?	YES ..... 1 → NO ..... 2 (e)	|___|___|___|___|	|___|___| MONTHS ..... 1 YEARS 2	|___|___| PER DAY ..... 1 PER WEEK 2
e. Urinary Obstruction Control Drugs. (Calcium Channel Blockers) (eg: Calan, Isoptin, Covera-HS, Varelen, Cardene, Adalat, Procardia, Cardura, Hytrin, Flomax,)	YES ..... 1 → NO ..... 2 (f)	|___|___|___|___|	|___|___| MONTHS 1 YEARS 2	|___|___| PER DAY .... 1 PER WEEK . 2
f. Viagra, Cialis, Levitra.	YES ..... 1 → NO 2 (C7)	|___|___|___|___|	|___|___| MONTHS 1 YEARS 2	|___|___| PER DAY .... 1 PER WEEK . 2

C2. Now I have some questions about supplements and other drugs some men take.

OTHER DRUGS AND SUPPLEMENTS	C5. Did you ever take (SUPPLEMENT)?	C6. In what year did you start to take (SUPPLEMENT)?	C7. How long did you take (SUPPLEMENT)?	C8. How often did you take (SUPPLEMENT) per day or per week?
a. DES (Diethyl stilbesterol)	YES……..1 → NO .....2 (b)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
b. Prostate Healthcare Drugs (ex: PC SPES, Saw Palmetto, Dayto, Homimex, Yoshimba, Damiana leave) Which one? _____	YES ..... 1 → NO ..... 2 (c)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
c. Lasix	YES ..... 1→ NO ..... 2(d)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
d. Lycopene	YES ..... 1 → NO ..... 2 (e)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
e. Selenium	YES ..... 1 →	|___|___|___|___|	|__|__| MONTHS….1	|__|__| PER DAY 1

	NO ..... 2 (f)		YEARS…….2	PER WEEK 2
f. Vitamin E	YES ..... 1 → NO ..... 2(g)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
g. Body Building or performance enhancing steroids.(DHEA, 19-Nor/androstenedione) Which one?_____	YES ..... 1 → NO ..... 2(h)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
h. Statins or Cholersterol lowering drugs (ex. Lipitor, Zocor, Mevacor) Which one?_____	YES ..... 1 → NO ..... 2 (i)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
i. Cox-2 Inhibitors (Celebrex, Vioxx, Bextra)	YES ..... 1 → NO ..... 2 (j)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
j.Multivitamin. Which one(s)? _____ _____	YES ..... 1 → NO ..... 2 ()	|___|___|___|___	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
k. Other Vitamins. Which one(s)? _____ _____	YES ..... 1 → NO ..... 2 ()	|___|___|___|___	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2

C3. Have you ever taken non-steroidal anti-inflammatory drugs (NSAIDs) such as Aspirin, Bufferin, Excedrin, Advil, Motrin, Nasproxsyn, and Ibuprofen (Tylenol is not an NSAID)?

( )$_0$ No (Skip to D) ( )$_1$ Occasionally (Skip to D) ( )$_2$ Weekly (Skip to D) ( )$_3$ Daily

C4. For what reason did you take NSAIDs?

( )$_0$ Headache ( )$_1$ Heart disease ( )$_2$ Stroke

( )$_3$ Arthritis ( )$_4$ Other_____(please specify)

C5. If you have taken NSAIDs on a daily basis, I would like to ask you about these periods during different times of your life. (Fill in table below)

Action	Period 1	Period 2	Period 3	Period 4	Period 5
a. In what year did you start taking these drugs?					
b. How many or how much did you take per day?	_____ ( )pills ( )mg	_____ ( )pills ( )mg	_____ ( )pills ( )mg	_____ ( )pills ( )mg	_____ ( )pills ( )mg
c.Which type or brand did you use?	_____	_____	_____	_____	_____
d. Did you continue to take this, stop or Δ your pattern for	( )$_0$ continued ( )$_1$ stopped ( )$_2$ pattern Δ	( )$_0$ continued ( )$_1$ stopped ( )$_2$ pattern Δ	( )$_0$ continued ( )$_1$ stopped ( )$_2$ pattern Δ	( )$_0$ continued ( )$_1$ stopped ( )$_2$ pattern Δ	( )$_0$ continued ( )$_1$ stopped ( )$_2$ pattern Δ

more than 6 months?					
e. Year you stopped taking NSAIDS or Δ your pattern for >6 months?	_____ If this is a Δ of pattern, ⇒C2a	_____ If this is a Δ of pattern, ⇒C3a	_____ If this is a Δ of pattern, ⇒C4a	_____ If this is a Δ of pattern, ⇒C5a	_____
f. Did you start NSAIDS again?	( )$_0$ no ⇒C6 ( )$_1$ yes ⇒C2a	( )$_0$ no ⇒C6 ( )$_1$ yes ⇒C2a	( )$_0$ no ⇒C6 ( )$_1$ yes ⇒C2a	( )$_0$ no ⇒C6 ( )$_1$ yes ⇒C2a	( )$_0$ no ( )$_1$ yes

C6. Have you taken any other prescription or non-prescription medications within the last year?

( )$_0$ No **(Skip to D)** ( )$_1$ Yes

C7. Which ones?

Name of Medication	Date began?	Date finished?	Reason for taking?	Notes

MEDICATIONS ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

D. SMOKING HISTORY

Now I have some questions about smoking.

D1. Have you ever smoked a total of 100 cigarettes or more in your lifetime?

( )$_0$ No **(Skip to E1)** ( )$_1$ Yes

D2. Did you ever smoke cigarettes regularly, at least one cigarette per day for six months or longer?

( )$_0$ No **(Skip to E1)** ( )$_1$ Yes

D3. How old were you when you first started smoking regularly? |___|___|

AGE STARTED

D4. Do you smoke cigarettes regularly now?

( )$_0$ No ( )$_1$ Yes (**Skip to D6)**

D5. How old were you when you stopped smoking regularly? |___|___|

AGE STOPPED

D6. In total, how many years have you smoked or did you smoke regularly (please subtract out years you did not smoke)?

|__|__|
YEARS

D7. Thinking about all the years when you smoked regularly, how many cigarettes did you usually smoke in a day?

|__|__|__|
CIGARETTES/DAY

D8. During your childhood, until you were 18, did anyone in your home smoke? (do not include this if smoking was done only outside the home). ( )$_0$ No (skip to D10) ( )$_1$ Yes

D9. How many people smoked in your home during your childhood? _____

D10. As an adult, does/did your spouse or partner or anyone else smoke in your home? (do not include this if smoking is/was done only outside the home). ( )$_0$ No ( )$_1$ Yes

D11. How many people smoked in your home during your adulthood? _____

D12. Do/Did you work in a place where co-workers smoked in your immediate area? ( )$_0$ No ( )$_1$ Yes

D13. For how many years were you working at a job where people smoked regularly in your immediate work area _____

SMOKING HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

E. ALCOHOL HISTORY

E1. Did you ever drink any alcohol beverages, such as beer, wine or hard liquor, on a regular basis, that is, at least once a week for 6 months or longer?

( )$_0$ No **(Skip to F1)** ( )$_1$ Yes

E2. How old were you when you started drinking regularly? |__|__|
AGE STARTED

E3. Do you still drink regularly now? ( )$_0$ No ( )$_1$ Yes **(Skip to E5)**

E4. How old were you when you stopped drinking regularly? |__|__|
AGE STOPPED

E5. In total, for how many years have you or did you drink regularly? Please subtract out the years when you didn't drink regularly.

|__|__|
YEARS

E6. On the average, after age 25, how many (ALCOHOLIC BEVERAGE) did you drink per week?	E7. How many years did you drink (ALCOHOLIC BEVERAGE) regularly?

DRINKS	YEARS
1. .....Cans or Bottles of Beer |___|___|___|	|___|___|
2. ..... Glasses of Wine |___|___|___|	|___|___|
4. ..... Shots of hard liquor |___|___|___|	|___|___|

ALCOHOL HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

F. OCCUPATIONAL HISTORY

We would like some information about the types of jobs you had **for the longest period of time.**

F1. What was the complete title of this job?_____

F2. Was this position a full-time or part-time job? (Full-time is 35 hours or more per week)
( )$_0$ Full-time
( )$_1$ Part-time

F3. What type of business or industry was this; that is what did this employer make or do? Please be as specific as possible._____

F4. What year did you begin this job and what year did you stop? ___/___ - ___/___
mo. yr mo. Yr

F5. What are/were your usual activities in this job?
_____

OCCUPATIONAL HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

G. BODY SIZE/ ANTHROPOMETRY

G1. How tall are you? |___|___| |___|___| or |___|___|
FT INCHES CM

DON'T KNOW----------988

G2. When you were about 8-9 years old, compared to other boys your age, were you ….?

Short	1
Somewhat short	2
Average height	3
Somewhat tall or	4
Tall?	5
DON'T KNOW	8

G3. When you were about 20-25 years old, compared to other men your age, were you ….?

Short	1
Somewhat short	2
Average height	3
Somewhat tall or	4
Tall?	5
DON'T KNOW	8

At what age did you reach your adult height? _____years

G4. After age 25, what has been your usual weight? |___|___|___| or |___|___|
LBS KG
DON'T KNOW 998

G5. Have you lost weight in the last 5 years? ( )$_0$ No ( )$_1$ Yes **(Skip to G8)**

G6. How much weight did you lose? |___|___|___| (IF LT 10 LBS GO TO G8)
LBS

G7. In the past 5 years, did you lose this weight without trying? ( )$_0$ No ( )$_1$ Yes

IN G8-G9, ASK EACH AGE GROUP ENDING WITH CURRENT AGE GROUP

	Age group				In the past year (prior to diagnosis)
	In 2nd to 4th grade	20-29 yrs old	40-49 yrs old	60-69 yrs old	
G8. When you were (AGE GROUP), compared with other males in the same age group were you ...?					
.....Very thin	 1	 1	 1	 1	 1
.....Somewhat thin	 2	 2	 2	 2	 2
..... Average	 3	 3	 3	 3	 3
.....Somewhat heavy	 4	 4	 4	 4	 4
..... Very heavy	 5	 5	 5	 5	 5
.....DON'T KNOW	 8	 8	 8	 8	 8
.....NOT APPLICABLE	 0	 0	 0	 0	 0
G9. What was your average weight at/in (AGE GROUP)?	|__|__|__| LBS	|__|__|__| LBS	|__|__|__| LBS	|__|__|__| LBS	|__|__|__| LBS
.....DON'T KNOW	998	998	998	998	998

G10. As an adult, what was your highest weight? |___|___|___| or |___|___|
LBS KG

G11. At what age did you first reach this highest weight? |___|___|
AGE

G12. For how many years or months were you at this highest weight? |___|___| MONTHS 1
YEARS 2

G13. When you gain weight, where on your body do you mainly tend to add the weight?
( )$_0$ don't gain weight
( )$_1$ around the waist and stomach
( )$_2$ around the hips and thighs
( )$_3$ around the chest and shoulders
( )$_4$ equally all over
( )$_5$ other (specify) _____

G14. Interviewer will ask: I would now like to measure your waist circumference (use standardized measurements- waist is belly button, hips are hip bone)

Waist circumference (cm)

First	Second		Difference	Tolerance	Third
|__|__|__|.|__|	|__|__|__|.|__|	|__|__|__|.|__|	2.0	|__|__|__|.|__|	

G15. Interviewer will ask: I would now like to measure your hip circumference.

Hip circumference (cm)

First	Second		Difference	Tolerance	Third
|__|__|__|.|__|	|__|__|__|.|__|	|__|__|__|.|__|	2.0	|__|__|__|.|__|	

G16. How would you describe your chest hair density? ( )$_0$ thick ( )$_1$ medium ( )$_2$ thin ( )$_3$ no hairs

G17. Have you experienced any permanent hair loss from your scalp since you were twenty years old? ( )$_0$ No ( )$_1$ Yes

G18. If yes, at what age did the hair loss begin? __ __ years

G19. Interviewer: Please indicate hair thickness ( )$_0$ thick ( )$_1$ medium ( )$_2$ thin ( )$_3$ no hairs

G20. Interviewer: Please indicate hair pattern on dome
( )$_0$ no evident loss
( )$_1$ some loss
( )$_2$ patterned baldness
( )$_3$ few hairs
( )$_4$ no hairs

Some Loss

Patterned Baldness

G21. Have you ever used any hair growth products? ( )$_0$ No ( )$_1$ Yes

G22. Are you using a wig or toupee? ( )$_0$ No ( )$_1$ Yes

BODY SIZE/ANTHROPOMETRY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

H. MEDICAL HISTORY

Now I am going to ask some questions about your health.

H1. Has a doctor ever told you that you had any of the following diseases? FOR EACH *YES* RESPONSE ASK I2. FOR EACH *NO* RESPONSE GO THE NEXT DISEASE				H2. IF YES Please tell me how old you were when the disease was (first) diagnosed.
	YES	NO		AGE
a.....Peptic ulcer	1	0	(b)	a. |___|___|
b. ..... Liver cirrhosis	1	0	(c)	b. |___|___|
c..... Other liver diseases	1	0	(d)	c. |___|___|
d. .....Hepatitis B	1	0	(e)	g. |___|___|
e.....Hepatitis C	1	0	(I3)	h. |___|___|

H3. Have you ever been told by a doctor that you have diabetes or sugar diabetes?
( )$_0$ No
( )$_1$ Yes

H4. At what age did your doctor first tell you this? __ __ years

H5. Are you now taking insulin?
( )$_0$ No **(Skip to H.8)**
( )$_1$ Yes

H6. At what age did you begin to take insulin? __ __ years

H7. For what reason do you take insulin? _____

H8. Are you now taking pills to lower you blood sugar? These are sometimes called oral agents or oral hypoglycemic agents?
( )$_0$ No **(Skip to I)** ( )$_1$ Yes

H9. At what age did you begin to take hypoglycemic agents? __ __ years

H10. For what reason do you take hypoglycemic agents? _____

MEDICAL HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

I. PROSTATE CANCER SCREENING HISTORY/UROLOGIC HEALTH

Now I'd like to ask you some questions about your urologic health.

Screening History

I1. Do you know the approximate date of your most recent examination (PSA test, DRE) for prostate cancer?
____/____/_____ _____ Don't remember (___ Never had examination (skip to I13)

I2. Was this examination performed by: _____your physician $_0$
_____a new physician who you did not know previously $_1$
_____in a free prostate cancer screening program $_2$

I3. Was the prostate exam done because you were experiencing any prostate-related symptoms (e.g., urinary control, pain)? ___yes$_1$ ____no$_0$ ____don't know$_8$

I4. Was your Digital Rectal Examination abnormal? ____yes$_1$ ____no$_0$ ____don't know$_8$

I5. Were you told that your PSA was elevated? ____yes$_1$ ____no$_0$ (skip to I8) ____don't know$_8$

I6. If so, what was your PSA value? _____(don't know=888)

I7. Did you follow up with further testing? ____yes$_1$ ____no$_0$

I8. Before this last exam , have you ever had an abnormal exam in the past (meaning that your doctor thought there was something that needed to be checked out further)? ____yes$_1$ ____no$_0$ ____don't know$_8$_____Never had exam before this last one$_9$.

I9. [IF YES] Have you had a biopsy previously? ____yes$_1$ ____no$_0$ ____don't know$_8$

a. Biopsy type	Diagnosis	Date	Hospital	Doctor
_____	_____	___/___/_____	_____	_____
_____	_____	___/___/_____	_____	_____
_____	_____	___/___/_____	_____	_____

I10. How often do you get checked out for prostate cancer?

_____every 3-6 months$_0$
_____annually$_1$
_____every 2 years$_2$
_____less often$_3$
_____don't know$_8$

I11. Approximately how many times would you say you have been checked for prostate cancer in your lifetime?

(*This would include the PSA and/or DRE)* _____(Don't know=888)

I12. Have you ever been screened in a free, mass screening program? ____yes$_1$ ____no$_0$

Urologic Health/History

I13. During a typical night, how many times do you wake up to urinate? *(For cases, please ask about a typical night during the 12 months prior to the prostate cancer diagnosis)*

( )$_0$ never (Skip to I15)
( )$_1$ once (Skip to I15)
( )$_2$ twice
( )$_3$ three times
( )$_4$ more than three times

I14. How old were you when you first began waking to urinate more than once a night on a regular basis?

___ ___ years

I15. Did a doctor ever tell you that you had:	Yes/No	How old were you when you were diagnosed?
a. an enlarged prostate or benign prostatic hypertrophy	( )$_0$ No ( )$_1$ Yes ( )$_8$ Don't know	
b. an inflamed prostate or prostatitis	( )$_0$ No ( )$_1$ Yes ( )$_8$ Don't know	
c. some other problem or disorder related to the urinary tract (specify) _____	( )$_0$ No ( )$_1$ Yes ( )$_8$ Don't know	
d. Some other problem or disorder related to the prostate (specify) _____	( )$_0$ No ( )$_1$ Yes ( )$_8$ Don't know	

I16. Have you ever had any prostate surgery?

( )$_0$ No **(Skip to I19)**
( )$_1$ Yes

I17. How many prostate surgeries have you had? _____

J18.	Year of surgery	Hospital name	City	State
a.				
b.				
c.				

I19. Were you ever treated by a doctor for a urinary tract infection since the age of 25?

( )$_0$ No
( )$_1$ Yes

I20. How old were you when your doctor first told you that you had a urinary tract infection?

___ ___ years

I21. How many times have you been diagnosed with a UTI?_____

I22. Have you had a vasectomy, that is a sterilization operation for men?

( )$_0$ No **(Skip to I24)**
( )$_1$ Yes

I23. How old were you when you had a vasectomy? __ __ years

I24. Were you circumcised? *Circumcision: The surgical removal of the foreskin of the penis.*

( )$_0$ No **(Skip to J)**
( )$_1$ Yes

I25. At what age were you circumcised?

( )$_1$ newborn
( )$_2$ other (specify in years) _____

PROSTATE HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

J. FAMILY MEDICAL HISTORY

J1. Has anyone in your family that is related to you by blood, ever been told he had Benign Prostatic Hyperplasia or an enlarged prostate? Include your sons, grandsons, father, paternal grandfather, maternal grandfather and brothers. ( )$_0$ No ( )$_1$ Yes

J2. If yes, at what age was it diagnosed?

Relative		Age at diagnosis (approximately) DK= 888
a	Brother(s) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
b	Father ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
c	Son (s) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
d	Maternal Grandfather ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
e	Paternal Grandfather ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
f	Other _____(specify) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	

J3. Has anyone in your family that is related to you by blood, ever been told he had prostate cancer? Include your sons, grandsons, father, paternal grandfather, maternal grandfather, brothers.

( )$_0$ No **(Skip to J5)** ( )$_1$ Yes

J4. If yes, at what age was it diagnosed?

Relative		Age at diagnosis (approximately) DK= 888
a	Brother(s) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
b	Father ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
c	Son (s) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
d	Maternal Grandfather ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
e	Paternal Grandfather ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
f	Other _____(specify) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	

J5. Has any member of your family that is related to you by blood ever been told that she had breast cancer? Including your daughter, mother, sister, grandmothers. ( )$_0$ No **(Skip to J7)**

( )$_1$ Yes

J6. If yes, at what age was it diagnosed?

Relative		Age at diagnosis (approximately) DK= 888
a	Daughter ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
b	Mother ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
c	Sister ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	

d	Maternal Grandmother ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
e	Paternal Grandmother ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
f	Other _____(specify) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	

J7. Have any members of your family that are related to you by blood ever been told that they had ovarian cancer? Please include your mother, daughter, and maternal and paternal grandmothers.

J8. If yes, at what age was it diagnosed?

Relative		Age at diagnosis (approximately) DK= 888
a	Daughter ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
b	Mother ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
c	Sister ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
d	Maternal Aunt ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
e	Paternal Grandmother ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
f	Other _____(specify) ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	

J9. Have any members of your family that are related to you by blood ever been told that they had endometrial cancer? Please include your mother, daughter, sisters and maternal and paternal grandmothers. ( )$_0$ No **(Skip to K)** ( )$_1$ Yes

J10. If yes, at what age was it diagnosed?

Relative		Age at diagnosis (approximately) DK= 888
a	Daughter ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
b	Mother ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
c	Sister(s) ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
d	Maternal Aunt ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
e	Paternal Grandmother ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
f	Other _____(specify) ( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	

FAMILY MEDICAL HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

K. PHYSICAL ACTIVITY/EXERCISE

Now, we are going to ask you about your levels of physical activity at different times in your life.

	a. Last year	b. Age 13-19	c. 20s	d. 30s	e. 40s	f. 50s+
K1. Did you participate in any routine physical activity for at least 20 minutes at a time that either made you sweat or increased your heart rate?	0 No 1 Yes	0 No 1 Yes	0 No 1 Yes	0 No 1 Yes	0 No 1 Yes	0 No 1 Yes
K2. What intensity level was your usual activity?	1 Moderate 2 Vigorous	1 Moderate 2 Vigorous	1Moderate 2 Vigorous	1 Moderate 2 Vigorous	1 Moderate 2 Vigorous	1 Moderate 2 Vigorous
K3. How often did you participate in this physical activity?	1 Less than 1x/week 2 1x/week 3 more than 1x/week	1 Less than 1x/week 2 1x/week 3 more than 1x/week	1 Less than 1x/week 2 1x/week 3 more than 1x/week	1 Less than 1x/week 2 1x/week 3 more than 1x/week	1 Less than 1x/week 2 1x/week 3 more than 1x/week	1 Less than 1x/week 2 1x/ week 3 more than 1x/week

PHYSICAL ACTIVITY ( )1 Very Good ( )2 Good ( )3 Fair ( )4 Poor

Section L (Sexual history) is self-administered, and the person will be given 20 min to complete this section.

SITE ID:______________________

L. SEXUAL HISTORY/HEALTH (self administered)

L1. At what age did you experience puberty (voice change, growth of pubic hair)? __ __ years

L2. How old were you when you first had sexual intercourse? __ __ years

	In your teens	In your 20's	In your 30's	In your 40's	In your 50's	In your 60's	In your 70's
L3.When you were (age group) with how many different partners did you have intercourse?	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ 40 or more	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ 40 or more	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ 40 or more	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ 40 or more	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ 40 or more	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ 40 or more	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ 40 or more
L4.If you think back to when you were (age group), and you think about the period of time in that decade when you had sexual intercourse, how often would you say you had sexual intercourse per month or per year?	________ times per ( ) month$_1$ ( ) year$_2$	________ times per ( ) month$_1$ ( ) year$_2$	________ times per ( ) month$_1$ ( ) year$_2$	________ times per ( ) month$_1$ ( ) year$_2$	________ times per ( ) month$_1$ ( ) year$_2$	________ times per ( ) month$_1$ ( ) year$_2$	________ times per ( ) month$_1$ ( ) year$_2$

L5. How many live-born children have you fathered? Do not include any stepchildren, foster children, or adopted children. __ __ __ **(If zero, skip to L7)**

L6. How old were you when your first child was born? __ __ years

L7. Have you ever tried to conceive a child for one year or more without success? ( )$_0$ No ( )$_1$ Yes

SITE ID:_____

L8. Did a doctor ever say that you had a problem that might be related to your difficulty in conceiving a child? If so, what was the problem? ( )$_0$ Low sperm count ( )$_1$ Low sperm motility ( )$_2$ Impotence ( )$_3$ Other _____(specify)

L9. Have you ever used condoms (rubbers)? ( )$_0$ No **(If No, skip to L13)** ( )$_1$ Yes

L10 Not counting the times that you were trying to conceive a child, how often did you use condoms? ( )$_0$ Rarely ( )$_1$ Sometimes ( )$_2$ Always

L11. Before one year ago, did you usually use condoms (rubbers)? ( )$_0$ No ( )$_1$ Yes

L12. Not counting the past year, for how many years did you use condoms (rubbers)?_____ YEARS

For the next question, please think about any sexually transmitted diseases that you may have contracted during your life.

L13.	Did a doctor ever tell you that you had:	Yes/No	How old were you when you were first diagnosed?	How many times altogether have you had the disease?
a.	Gonorrhea	( )$_0$ No ( )$_1$ Yes		
b.	Syphilis	( )$_0$ No ( )$_1$ Yes		
c.	Genital Warts	( )$_0$ No ( )$_1$ Yes		
d.	Genital Herpes	( )$_0$ No ( )$_1$ Yes		
e.	Other sexually transmitted disease _____(specify)	( )$_0$ No ( )$_1$ Yes		
f.	Other sexually transmitted disease _____(specify)	( )$_0$ No ( )$_1$ Yes		

This completes our interview. I would like to now take the samples and I want to thank you very much for the time you have spent in answering my questions today.

May we contact you again later if we need to clarify any of the information you have provided?
( )0 No ( )1 Yes

Time ended: ___ ___ : ___ ___ ( )1 AM
( )2 PM

M. ADMINISTRATIVE INFORMATION

M1. Date form completed ___ ___ / ___ ___ / ___ ___ ___ ___

M2. Name of interviewer _____/_____/_____

M3. Interviewer ID Number: ___ ___

M4. Interviewer's Signature: _____

N. INTERVIEWER REMARKS

N1. Interview was conducted:
( )1 In the clinic_____
( )2 General Clinical Research Center
( )3 Over the phone
( )4 Other (specify)_____

N2. Respondent's cooperation was:
( )1 Very good
( )2 Good
( )3 Fair
( )4 Poor

N3. The overall quality of the interview was:
( )1 Very good
( )2 Good
( )3 Fair
( )4 Poor

N4. Did any of the following occur during the interview?

a.	R did not know enough information regarding the topics.	( )0 No	( )1 Yes
b.	R did not want to be more specific.	( )0 No	( )1 Yes
c.	R did not understand or speak English well.	( )0 No	( )1 Yes
d.	R was upset or depressed.	( )0 No	( )1 Yes
e.	R had poor hearing or speech.	( )0 No	( )1 Yes
f.	R was confused by frequent interruptions.	( )0 No	( )1 Yes
g.	R was emotionally unstable.	( )0 No	( )1 Yes

h.	Others helped with the answers.	( )0 No ( )1 Yes
i.	R required a lot of probing	( )0 No ( )1 Yes
j.	Patient was reserved	( )0 No ( )1 Yes
k.	R was physically ill	( )0 No ( )1 Yes
l.	Other, (specify) _____	( )0 No ( )1 Yes

N5. Comments/Remarks:

_____
_____
_____
_____
_____
_____
_____

NATIONAL INSTITUTES OF HEALTH

Diet History Questionnaire

GENERAL INSTRUCTIONS

- **Answer each question as best you can. Estimate if you are not sure. A guess is better than leaving a blank.**
- **Use only a black ball-point pen. Do not use a pencil or felt-tip pen. Do not fold, staple, or tear the pages.**
- **Put an X in the box next to your answer.**
- **If you make any changes, cross out the incorrect answer and put an X in the box next to the correct answer. Also draw a circle around the correct answer.**
- **If you mark NEVER, NO, or DON'T KNOW for a question, please follow any arrows or instructions that direct you to the next question.**

BEFORE TURNING THE PAGE, PLEASE COMPLETE THE FOLLOWING QUESTIONS.

Today's date:

MONTH	DAY		YEAR
☐ Jan	|___|___|		☐ 2002
☐ Feb	☐0	☐0	☐ 2003
☐ Mar	☐1	☐1	☐ 2004
☐ Apr	☐2	☐2	☐ 2005
☐ May	☐3	☐3	☐ 2006
☐ Jun		☐4	
☐ Jul		☐5	
☐ Aug		☐6	
☐ Sep		☐7	
☐ Oct		☐8	
☐ Nov		☐9	
☐ Dec			

In what month were you born?

☐ Jan
☐ Feb
☐ Mar
☐ Apr
☐ May
☐ Jun
☐ Jul
☐ Aug
☐ Sep
☐ Oct
☐ Nov
☐ Dec

In what year were you born?

19 |___|___|

☐0	☐0
☐1	☐1
☐2	☐2
☐3	☐3
☐4	☐4
☐5	☐5
☐6	☐6
☐7	☐7
☐8	☐8
☐9	☐9

Are you male or female?

☐Male
☐Female

BAR CODE LABEL OR SUBJECT ID HERE

|___|___|___|___|___|___|___|___|

1. Over the past 12 months, how often did you drink **tomato juice** or **vegetable juice**?

☐ NEVER (GO TO QUESTION 2)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

1a. Each time you drank **tomato juice** or **vegetable juice**, how much did you usually drink?

☐ Less than ¾ cup (6 ounces)
☐ ¾ to 1¼ cups (6 to 10 ounces)
☐ More than 1¼ cups (10 ounces)

2. Over the past 12 months, how often did you drink **orange juice** or **grapefruit juice?**

☐ NEVER (GO TO QUESTION 3)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

2a. Each time you drank **orange juice** or **grapefruit juice**, how much did you usually drink?

☐ Less than ¾ cup (6 ounces)
☐ ¾ to 1¼ cups (6 to 10 ounces)
☐ More than 1¼ cups (10 ounces)

3. Over the past 12 months, how often did you drink **other 100% fruit juice** or **100% fruit juice mixtures** (such as apple, grape, pineapple, or others)?

☐ NEVER (GO TO QUESTION 4)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

3a. Each time you drank **other fruit juice** or **fruit juice mixtures**, how much did you usually drink?

☐ Less than ¾ cup (6 ounces)
☐ ¾ to 1½ cups (6 to 12 ounces)
☐ More than 1½ cups (12 ounces)

Question 4 appears in the next column

Over the past 12 months…

4. How often did you drink other **fruit drinks** (such as cranberry cocktail, Hi-C, lemonade, or Kool-Aid, diet or regular)?

☐ NEVER (GO TO QUESTION 5)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

4a. Each time you drank **fruit drinks**, how much did you usually drink?

☐ Less than 1 cup (8 ounces)
☐ 1 to 2 cups (8 to 16 ounces)
☐ More than 2 cups (16 ounces)

4b. How often were your fruit drinks **diet** or **sugar-free drinks**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

5. How often did you drink **milk as a beverage** (NOT in coffee, NOT in cereal)? (Please include chocolate milk and hot chocolate.)

☐ NEVER (GO TO QUESTION 6)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

5a. Each time you drank **milk as a beverage**, how much did you usually drink?

☐ Less than 1 cup (8 ounces)
☐ 1 to 1½ cups (8 to 12 ounces)
☐ More than 1½ cups (12 ounces)

5b. What kind of **milk** did you usually drink?

☐ Whole milk
☐ 2% fat milk
☐ 1 % fat milk
☐ Skim, nonfat, or ½% fat milk
☐ Soy milk
☐ Rice milk
☐ Other

Question 6 appears on the next page

Over the <u>past 12 months</u>…

6. How often did you drink **meal replacement, energy, or high-protein beverages** such as Instant Breakfast, Ensure, Slimfast, Sustacal or others?

☐ NEVER (GO TO QUESTION 7)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

6a. Each time you drank **meal replacement beverages**, how much did you usually drink?

☐ Less than 1 cup (8 ounces)
☐ 1 to 1½ cups (8 to 12 ounces)
☐ More than 1½ cups (12 ounces)

7. Over the past 12 months, did you drink **soft drinks, soda,** or **pop**?

☐ NO (GO TO QUESTION 8)

☐ YES

7a. How often did you drink **soft drinks, soda,** or **pop IN THE SUMMER**?

☐ NEVER

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

7b. How often did you drink **soft drinks, soda,** or **pop DURING THE REST OF THE YEAR**?

☐ NEVER

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

7c. Each time you drank **soft drinks, soda,** or **pop,** how much did you usually drink?

☐ Less than 12 ounces or less than 1 can or bottle
☐ 12 to 16 ounces or 1 can or bottle
☐ More than 16 ounces or more than 1 can or bottle

7d. How often were these soft drinks, soda, or pop **diet** or **sugar-free**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

7e. How often were these soft drinks, soda, or pop **caffeine-free**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

8. Over the <u>past 12 months</u>, did you drink **beer**?

☐ NO (GO TO QUESTION 9)

☐ YES

8a. How often did you drink **beer IN THE SUMMER**?

☐ NEVER

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

8b. How often did you drink **beer DURING THE REST OF THE YEAR**?

☐ NEVER

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

8c. Each time you drank **beer**, how much did you usually drink?

☐ Less than a 12-ounce can or bottle
☐ 1 to 3 12-ounce cans or bottles
☐ More than 3 12-ounce cans or bottles

Question 8 appears in the next column

Question 9 appears on the next page

Over the <u>past 12 months</u>…

9. How often did you drink **wine** or **wine coolers**?

☐ NEVER (GO TO QUESTION 10)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

9a. Each time you drank **wine** or **wine coolers,** how much did you usually drink?

☐ Less than 5 ounces or less than 1 glass
☐ 5 to 12 ounces or 1 to 2 glasses
☐ More than 12 ounces or more than 2 glasses

10. How often did you drink **liquor** or **mixed drinks**?

☐ NEVER (GO TO QUESTION 11)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

10a. Each time you drank **liquor** or **mixed drinks,** how much did you usually drink?

☐ Less than 1 shot of liquor
☐ 1 to 3 shots of liquor
☐ More than 3 shots of liquor

11. Over the <u>past 12 months</u>, did you eat **oatmeal, grits,** or **other cooked cereal?**

☐ NO (GO TO QUESTION 12)

☐ YES

11a. How often did you eat **oatmeal, grits,** or **other cooked cereal IN THE WINTER**?

☐ NEVER

☐ 1–6 times per winter
☐ 7–11 times per winter
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

11b. How often did you eat **oatmeal, grits,** or **other cooked cereal DURING THE REST OF THE YEAR**?

☐ NEVER

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

11c. Each time you ate **oatmeal, grits,** or **other cooked cereal,** how much did you usually eat?

☐ Less than ¾ cup
☐ ¾ to 1¼ cups
☐ More than 1¼ cups

12. How often did you eat **cold cereal**?

☐ NEVER (GO TO QUESTION 13)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

12a. Each time you ate **cold cereal**, how much did you usually eat?

☐ Less than 1 cup
☐ 1 to 2½ cups
☐ More than 2½ cups

12b. How often was the cold cereal you ate **Total, Product 19,** or **Right Start**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

12c. How often was the cold cereal you ate **All Bran, Fiber One, 100% Bran,** or **Bran Buds?**

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 12 appears in the next column

Question 13 appears on the next page

Over the <u>past 12 months</u>…

12d. How often was the cold cereal you ate **some other bran** or **fiber cereal** (such as Cheerios, Shredded Wheat, Raisin Bran, Bran Flakes, Grape-Nuts, Granola, Wheaties, or Healthy Choice)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

12e. How often was the cold cereal you ate any **other type of cold cereal** (such as Corn Flakes, Rice Krispies, Frosted Flakes, Special K, Froot Loops, Cap'n Crunch, or others)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

12f. Was **milk** added to your cold cereal?

- ☐ NO (GO TO QUESTION 13)
- ☐ YES

12g. What kind of **milk** was usually added?

- ☐ Whole milk
- ☐ 2% fat milk
- ☐ 1% fat milk
- ☐ Skim, nonfat, or ½% fat milk
- ☐ Soy milk
- ☐ Rice milk
- ☐ Other

12h. Each time **milk was added to your cold cereal**, how much was usually added?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

13. How often did you eat **applesauce**?

- ☐ NEVER (GO TO QUESTION 14)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

Question 14 appears in the next column

13a. Each time you ate **applesauce**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

14. How often did you eat **apples**?

- ☐ NEVER (GO TO QUESTION 15)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

14a. Each time you ate **apples**, how many did you usually eat?

- ☐ Less than 1 apple
- ☐ 1 apple
- ☐ More than 1 apple

15. How often did you eat **pears** (fresh, canned, or frozen)?

- ☐ NEVER (GO TO QUESTION 16)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

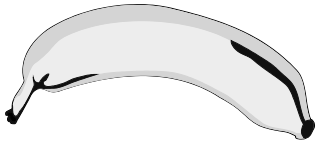
15a. Each time you ate **pears**, how many did you usually eat?

- ☐ Less than 1 pear
- ☐ 1 pear
- ☐ More than 1 pear

16. How often did you eat **bananas?**

- ☐ NEVER (GO TO QUESTION 17)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day



Question 17 appears on the next page

Over the <u>past 12 months</u>…

16a. Each time you ate **bananas**, how many did you usually eat?

- ☐ Less than 1 banana
- ☐ 1 banana
- ☐ More than 1 banana

17. How often did you eat **dried fruit**, such as prunes or raisins (not including dried apricots)?

- ☐ NEVER (GO TO QUESTION 18)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

17a. Each time you ate **dried fruit**, how much did you usually eat (not including dried apricots)?

- ☐ Less than 2 tablespoons
- ☐ 2 to 5 tablespoons
- ☐ More than 5 tablespoons

18. Over the <u>past 12 months</u>, did you eat **peaches, nectarines,** or **plums**?

- ☐ NO (GO TO QUESTION 19)
- ☐ YES

18a. How often did you eat **fresh peaches**, **nectarines**, or **plums WHEN IN SEASON**?

- ☐ NEVER

- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

18b. How often did you eat **peaches, nectarines,** or **plums** (fresh, canned, or frozen) **DURING THE REST OF THE YEAR**?

- ☐ NEVER

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

18c. Each time you ate **peaches, nectarines,** or **plums,** how much did you usually eat?

- ☐ Less than 1 fruit or less than ½ cup
- ☐ 1 to 2 fruits or ½ to ¾ cup
- ☐ More than 2 fruits or more than ¾ cup

19. How often did you eat **grapes**?

- ☐ NEVER (GO TO QUESTION 20)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

19a. Each time you ate **grapes**, how much did you usually eat?

- ☐ Less than ½ cup or less than 10 grapes
- ☐ ½ to 1 cup or 10 to 30 grapes
- ☐ More than 1 cup or more than 30 grapes

20. Over the <u>past 12 months</u>, did you eat **cantaloupe**?

- ☐ NO (GO TO QUESTION 21)
- ☐ YES

20a. How often did you eat **fresh cantaloupe WHEN IN SEASON**?

- ☐ NEVER

- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

20b. How often did you eat **fresh** or **frozen cantaloupe DURING THE REST OF THE YEAR**?

- ☐ NEVER

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 19 appears in the next column

Question 21 appears on the next page

Over the <u>past 12 months</u>…

20c. Each time you ate **cantaloupe**, how much did you usually eat?

- ☐ Less than ¼ melon or less than ½ cup
- ☐ ¼ melon or ½ to 1 cup
- ☐ More than ¼ melon or more than 1 cup

21. Over the <u>past 12 months</u>, did you eat **melon, other than cantaloupe** (such as watermelon or honeydew)?

- ☐ NO (GO TO QUESTION 22)
- ☐ YES

21a. How often did you eat **fresh melon, other than cantaloupe** (such as watermelon or honeydew) **WHEN IN SEASON**?

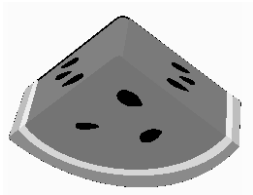
- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

21b. How often did you eat **fresh** or **frozen melon, other than cantaloupe** (such as watermelon or honeydew) **DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

21c. Each time you ate **melon other than cantaloupe**, how much did you usually eat?

- ☐ Less than ½ cup or 1 small wedge
- ☐ ½ to 2 cups or 1 medium wedge
- ☐ More than 2 cups or 1 large wedge



Question 22 appears in the next column

22. Over the <u>past 12 months</u>, did you eat **strawberries**?

- ☐ NO (GO TO QUESTION 23)
- ☐ YES

22a. How often did you eat **fresh strawberries WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

22b. How often did you eat **fresh** or **frozen strawberries DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

22c. Each time you ate **strawberries**, how much did you usually eat?

- ☐ Less than ¼ cup or less than 3 berries
- ☐ ¼ to ¾ cup or 3 to 8 berries
- ☐ More than ¾ cup or more than 8 berries

23. Over the <u>past 12 months</u>, did you eat **oranges, tangerines,** or **tangelos**?

- ☐ NO (GO TO QUESTION 24)
- ☐ YES

23a. How often did you eat **fresh oranges, tangerines,** or **tangelos WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 24 appears on the next page

Over the <u>past 12 months</u>…

23b. How often did you eat **oranges, tangerines,** or **tangelos** (fresh or canned) **DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

23c. Each time you ate **oranges, tangerines,** or **tangelos,** how many did you usually eat?

- ☐ Less than 1 fruit
- ☐ 1 fruit
- ☐ More than 1 fruit

24. Over the <u>past 12 months</u>, did you eat **grapefruit**?

- ☐ NO (GO TO QUESTION 25)
- ☐ YES

24a. How often did you eat **fresh grapefruit WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

24b. How often did you eat **grapefruit** (fresh or canned) **DURING THE REST OF THE YEAR?**

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

24c. Each time you ate **grapefruit**, how much did you usually eat?

- ☐ Less than ½ grapefruit
- ☐ ½ grapefruit
- ☐ More than ½ grapefruit

25. How often did you eat **other kinds of fruit**?

- ☐ NEVER (GO TO QUESTION 26)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

25a. Each time you ate **other kinds of fruit**, how much did you usually eat?

- ☐ Less than ¼ cup
- ☐ ¼ to ¾ cup
- ☐ More than ¾ cup

26. How often did you eat **COOKED greens** (such as spinach, turnip, collard, mustard, chard, or kale)?

- ☐ NEVER (GO TO QUESTION 27)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

26a. Each time you ate **COOKED greens**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

27. How often did you eat **RAW greens** (such as spinach, turnip, collard, mustard, chard, or kale)? (*We will ask about lettuce later.)*

- ☐ NEVER (GO TO QUESTION 28)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

27a. Each time you ate **RAW greens**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

Question 25 appears in the next column

Question 28 appears on the next page

Over the <u>past 12 months</u>…

28. How often did you eat **coleslaw**?

☐ NEVER (GO TO QUESTION 29)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

28a. Each time you ate **coleslaw**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to ¾ cup
☐ More than ¾ cup

29. How often did you eat **sauerkraut** or **cabbage** (other than coleslaw)?

☐ NEVER (GO TO QUESTION 30)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

29a. Each time you ate **sauerkraut** or **cabbage**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to 1 cup
☐ More than 1 cup

30. How often did you eat **carrots** (fresh, canned, or frozen)?

☐ NEVER (GO TO QUESTION 31)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

30a. Each time you ate **carrots**, how much did you usually eat?

☐ Less than ¼ cup or less than 2 baby carrots
☐ ¼ to ½ cup or 2 to 5 baby carrots
☐ More than ½ cup or more than 5 baby carrots

Question 31 appears in the next column

31. How often did you eat **string beans** or **green beans** (fresh, canned, or frozen)?

☐ NEVER (GO TO QUESTION 32)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

31a. Each time you ate **string beans** or **green beans**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 cup
☐ More than 1 cup

32. How often did you eat **peas** (fresh, canned, or frozen)?

☐ NEVER (GO TO QUESTION 33)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

32a. Each time you ate **peas**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to ¾ cup
☐ More than ¾ cup

33. Over the <u>past 12 months</u>, did you eat **corn**?

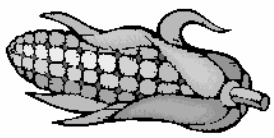
☐ NO (GO TO QUESTION 34)

☐ YES

33a. How often did you eat **fresh corn WHEN IN SEASON**?

☐ NEVER

☐ 1–6 times per season
☐ 7–11 times per season
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day



Question 34 appears on the next page

Over the <u>past 12 months</u>…

33b. How often did you eat **corn** (fresh, canned, or frozen) **DURING THE REST OF THE YEAR**?

☐ NEVER

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

33c. Each time you ate **corn**, how much did you usually eat?

☐ Less than 1 ear or less than ½ cup
☐ 1 ear or ½ to 1 cup
☐ More than 1 ear or more than 1 cup

34. Over the <u>past 12 months</u>, how often did you eat **broccoli** (fresh or frozen)?

☐ NEVER (GO TO QUESTION 35)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

34a. Each time you ate **broccoli**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to 1 cup
☐ More than 1 cup

35. How often did you eat **cauliflower** or **Brussels sprouts** (fresh or frozen)?

☐ NEVER (GO TO QUESTION 36)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

35a. Each time you ate **cauliflower** or **Brussels sprouts**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to ½ cup
☐ More than ½ cup

Question 36 appears in the next column

36. How often did you eat mixed **vegetables**?

☐ NEVER (GO TO QUESTION 37)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

36a. Each time you ate **mixed vegetables**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 cup
☐ More than 1 cup

37. How often did you eat **onions**?

☐ NEVER (GO TO QUESTION 38)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

37a. Each time you ate **onions**, how much did you usually eat?

☐ Less than 1 slice or less than 1 tablespoon
☐ 1 slice or 1 to 4 tablespoons
☐ More than 1 slice or more than 4 tablespoons

38. Now think about all the **cooked vegetables** you ate in the <u>past 12 months</u> and how they were prepared. How often were your vegetables **COOKED WITH** some sort of **fat**, including oil spray? *(Please do not include potatoes.)*

☐ NEVER (GO TO QUESTION 39)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

Question 39 appears on the next page

Over the <u>past 12 months</u>…

38a. Which fats were usually added to your vegetables **DURING COOKING**? *(Please do not include potatoes. **Mark all that apply.**)*

- ☐ Margarine (including low-fat)
- ☐ Butter (including low-fat)
- ☐ Lard, fatback, or bacon fat
- ☐ Olive oil
- ☐ Corn oil
- ☐ Canola or rapeseed oil
- ☐ Oil spray, such as Pam or others
- ☐ Other kinds of oils
- ☐ None of the above

39. Now, thinking again about all the **cooked vegetables** you ate in the <u>past 12 months</u>, how often was some sort of fat, sauce, or dressing added **AFTER COOKING OR AT THE TABLE**? *(Please do not include potatoes.)*

- ☐ NEVER (GO TO QUESTION 40)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1–2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 times per day
- ☐ 3 or more times per day

39a. Which fats, sauces, or dressings were usually added **AFTER COOKING OR AT THE TABLE**? *(Please do not include potatoes. **Mark all that apply.**)*

- ☐ Margarine (including low-fat)
- ☐ Butter (including low-fat)
- ☐ Lard, fatback, or bacon fat
- ☐ Salad dressing
- ☐ Cheese sauce
- ☐ White sauce
- ☐ Other

39b. If margarine, butter, lard, fatback, or bacon fat was added to your cooked vegetables **AFTER COOKING OR AT THE TABLE**, how much did you usually add?

- ☐ Did not usually add these
- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

39c. If salad dressing, cheese sauce, or white sauce was added to your cooked vegetables **AFTER COOKING OR AT THE TABLE**, how much did you usually add?

- ☐ Did not usually add these
- ☐ Less than 1 tablespoon
- ☐ 1 to 3 tablespoons
- ☐ More than 3 tablespoons

Question 40 appears in the next column

40. Over the <u>past 12 months</u>, how often did you eat **sweet peppers** (green, red, or yellow)?

- ☐ NEVER (GO TO QUESTION 41)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

40a. Each time you ate **sweet peppers**, how much did you usually eat?

- ☐ Less than ⅛ pepper
- ☐ ⅛ to ¼ pepper
- ☐ More than ¼ pepper

41. Over the <u>past 12 months</u>, did you eat **fresh tomatoes** (including those in salads)?

- ☐ NO (GO TO QUESTION 42)
- ☐ YES

41a. How often did you eat **fresh tomatoes** (including those in salads) **WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

41b. How often did you eat **fresh tomatoes** (including those in salads) **DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

41c. Each time you ate **fresh tomatoes**, how much did you usually eat?

- ☐ Less than ¼ tomato
- ☐ ¼ to ½ tomato
- ☐ More than ½ tomato

Question 42 appears on the next page

Over the <u>past 12 months</u>…

42. How often did you eat **lettuce salads** (with or without other vegetables)?

☐ NEVER (GO TO QUESTION 43)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

42a. Each time you ate **lettuce salads**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to 1 /4 cups
☐ More than 1 /4 cups

43. How often did you eat **salad dressing** (including low-fat) on salads?

☐ NEVER (GO TO QUESTION 44)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

43a. Each time you ate **salad dressing** on salads, how much did you usually eat?

☐ Less than 2 tablespoons
☐ 2 to 4 tablespoons
☐ More than 4 tablespoons

44. How often did you eat **sweet potatoes** or **yams**?

☐ NEVER (GO TO QUESTION 45)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

44a. Each time you ate **sweet potatoes** or **yams**, how much did you usually eat?

☐ 1 small potato or less than ¼ cup
☐ 1 medium potato or ¼ to ¾ cup
☐ 1 large potato or more than ¾ cup

45. How often did you eat **French fries, home fries, hash browned potatoes**, or **tater tots**?

☐ NEVER (GO TO QUESTION 46)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

45a. Each time you ate **French fries**, **home fries**, **hash browned potatoes**, or **tater tots** how much did you usually eat?

☐ Less than 10 fries or less than ½ cup
☐ 10 to 25 fries or ½ to 1 cup
☐ More than 25 fries or more than 1 cup

46. How often did you eat **potato salad**?

☐ NEVER (GO TO QUESTION 47)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

46a. Each time you ate **potato salad**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 cup
☐ More than 1 cup

47. How often did you eat **baked, boiled,** or **mashed potatoes**?

☐ NEVER (GO TO QUESTION 48)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

47a. Each time you ate **baked, boiled,** or **mashed potatoes**, how much did you usually eat?

☐ 1 small potato or less than ½ cup
☐ 1 medium potato or ½ to 1 cup
☐ 1 large potato or more than 1 cup

Question 45 appears in the next column

Question 48 appears on the next page

Over the <u>past 12 months</u>…

47b. How often was **sour cream** (including low-fat) added to your potatoes, **EITHER IN COOKING OR AT THE TABLE**?

- ☐ Almost never or never (GO TO QUESTION 47d)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

47c. Each time **sour cream** was added to your potatoes, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 3 tablespoons
- ☐ More than 3 tablespoons

47d. How often was **margarine** (including low-fat) added to your potatoes, **EITHER IN COOKING OR AT THE TABLE**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

47e. How often was **butter** (including low-fat) added to your potatoes, **EITHER IN COOKING OR AT THE TABLE**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

47f. Each time **margarine** or **butter** was added to your potatoes, how much was usually added?

- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

47g. How often was cheese or cheese sauce added to your potatoes, EITHER IN COOKING OR AT THE TABLE?

- ☐ Almost never or never (GO TO QUESTION 48)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

47h. Each time **cheese** or **cheese sauce** was added to your potatoes, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 3 tablespoons
- ☐ More than 3 tablespoons

48. How often did you eat **salsa**?

- ☐ NEVER (GO TO QUESTION 49)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

48a. Each time you ate **salsa**, how much did you usually eat?

- ☐ Less than 1 tablespoon
- ☐ 1 to 5 tablespoons
- ☐ More than 5 tablespoons

49. How often did you eat **catsup**?

- ☐ NEVER (GO TO QUESTION 50)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

49a. Each time you ate **catsup**, how much did you usually eat?

- ☐ Less than 1 teaspoon
- ☐ 1 to 6 teaspoons
- ☐ More than 6 teaspoons

50. How often did you eat **stuffing, dressing,** or **dumplings**?

- ☐ NEVER (GO TO QUESTION 51)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

50a. Each time you ate **stuffing, dressing,** or **dumplings**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

Question 48 appears in the next column

Question 51 appears on the next page

Over the <u>past 12 months</u>…

51. How often did you eat **chili**?

☐ NEVER (GO TO QUESTION 52)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

51a. Each time you ate **chili**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 3/4 cups
☐ More than 1 3/4 cups

52. How often did you eat **Mexican foods** (such as tacos, tostados, burritos, tamales, fajitas, enchiladas, quesadillas, and chimichangas)?

☐ NEVER (GO TO QUESTION 53)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

52a. Each time you ate **Mexican foods**, how much did you usually eat?

☐ Less than 1 taco, burrito, etc.
☐ 1 to 2 tacos, burritos, etc.
☐ More than 2 tacos, burritos, etc.

53. How often did you eat **cooked dried beans** (such as baked beans, pintos, kidney, blackeyed peas, lima, lentils, soybeans, or refried beans)? *(Please don't include bean soups or chili.)*

☐ NEVER (GO TO QUESTION 54)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

53a. Each time you ate **beans**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 cup
☐ More than 1 cup

53b. How often were the beans you ate **refried beans, beans prepared with any type of fat**, or **with meat added**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

54. How often did you eat **other kinds of vegetables**?

☐ NEVER (GO TO QUESTION 55)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

54a. Each time you ate **other kinds of vegetables**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to ½ cup
☐ More than ½ cup

55. How often did you eat **rice** or **other cooked grains** (such as bulgur, cracked wheat, or millet)?

☐ NEVER (GO TO QUESTION 56)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

55a. Each time you ate **rice** or **other cooked grains**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 /2 cups
☐ More than 1 /2 cups

55b. How often was **butter, margarine**, or **oil** added to your rice **IN COOKING OR AT THE TABLE**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 54 appears in the next column

Question 56 appears on the next page

Over the <u>past 12 months</u>…

56. How often did you eat **pancakes, waffles,** or **French toast**?

☐ NEVER (GO TO QUESTION 57)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

56a. Each time you ate **pancakes, waffles,** or **French toast**, how much did you usually eat?

☐ Less than 1 medium piece
☐ 1 to 3 medium pieces
☐ More than 3 medium pieces

56b. How often was **margarine** (including low-fat) added to your pancakes, waffles, or French toast **AFTER COOKING OR AT THE TABLE**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

56c. How often was **butter** (including low-fat) added to your pancakes, waffles, or French toast **AFTER COOKING OR AT THE TABLE**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

56d. Each time **margarine** or **butter** was added to your pancakes, waffles, or French toast, how much was usually added?

☐ Never added
☐ Less than 1 teaspoon
☐ 1 to 3 teaspoons
☐ More than 3 teaspoons

56e. How often was **syrup** added to your pancakes, waffles, or French toast?

☐ Almost never or never (GO TO QUESTION 57)
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 57 appears in the next column

56f. Each time **syrup** was added to your pancakes, waffles, or French toast, how much was usually added?

☐ Less than 1 tablespoon
☐ 1 to 4 tablespoons
☐ More than 4 tablespoons

57. How often did you eat **lasagna, stuffed shells, stuffed manicotti, ravioli,** or **tortellini**? *(Please do not include spaghetti or other pasta.)*

☐ NEVER (GO TO QUESTION 58)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

57a. Each time you ate **lasagna, stuffed shells, stuffed manicotti, ravioli,** or **tortellini**, how much did you usually eat?

☐ Less than 1 cup
☐ 1 to 2 cups
☐ More than 2 cups

58. How often did you eat **macaroni and cheese**?

☐ NEVER (GO TO QUESTION 59)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

58a. Each time you ate **macaroni and cheese**, how much did you usually eat?

☐ Less than 1 cup
☐ 1 to 1 /2 cups
☐ More than 1 /2 cups

59. How often did you eat **pasta salad** or **macaroni salad**?

☐ NEVER (GO TO QUESTION 60)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

Question 60 appears on the next page

Over the <u>past 12 months</u>…

59a. Each time you ate **pasta salad** or **macaroni salad**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

60. Other than the pastas listed in Questions 57, 58, and 59, how often did you eat **pasta, spaghetti**, or **other noodles**?

- ☐ NEVER (GO TO QUESTION 61)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

60a. Each time you ate **pasta, spaghetti**, **or other noodles**, how much did you usually eat?

- ☐ Less than 1 cup
- ☐ 1 to 3 cups
- ☐ More than 3 cups

60b. How often did you eat your pasta, spaghetti, or other noodles with **tomato sauce** or **spaghetti sauce made WITH meat**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

60c. How often did you eat your pasta, spaghetti, or other noodles with **tomato sauce** or **spaghetti sauce made WITHOUT meat**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

60d. How often did you eat your pasta, spaghetti, or other noodles with **margarine, butter, oil,** or **cream sauce**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 61 appears in the next column

61. How often did you eat **bagels** or **English muffins**?

- ☐ NEVER (GO TO INTRODUCTION TO QUESTION 62)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

61a. Each time you ate **bagels** or **English muffins**, how many did you usually eat?

- ☐ Less than 1 bagel or English muffin
- ☐ 1 bagel or English muffin
- ☐ More than 1 bagel or English muffin

61b. How often was **margarine** (including low-fat) added to your bagels or English muffins?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

61c. How often was **butter** (including low-fat) added to your bagels or English muffins?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

61d. Each time **margarine** or **butter** was added to your bagels or English muffins, how much was usually added?

- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 2 teaspoons
- ☐ More than 2 teaspoons

61e. How often was **cream cheese** (including low-fat) spread on your bagels or English muffins?

- ☐ Almost never or never (GO TO INTRODUCTION TO QUESTION 62)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Introduction to Question 62 appears on the next page

Over the <u>past 12 months</u>…

61f. Each time **cream cheese** was added to your bagels or English muffins, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 2 tablespoons
- ☐ More than 2 tablespoons

The next questions ask about your intake of breads other than bagels or English muffins. First, we will ask about bread you ate as part of sandwiches only. Then we will ask about all other bread you ate.

62. How often did you eat **breads** or **rolls AS PART OF SANDWICHES** (including burger and hot dog rolls)?

- ☐ NEVER (GO TO QUESTION 63)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

62a. Each time you ate **breads** or **rolls AS PART OF SANDWICHES**, how many did you usually eat?

- ☐ 1 slice or ½ roll
- ☐ 2 slices or 1 roll
- ☐ More than 2 slices or more than 1 roll

62b. How often were the breads or rolls that you used for your sandwiches **white bread** (including burger and hot dog rolls)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

62c. How often was **mayonnaise** or **mayonnaise-type dressing** (including low-fat) added to your sandwich bread or rolls?

- ☐ Almost never or never (GO TO QUESTION 62e)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 62e appears in the next column

Question 63 appears in the next column

62d. Each time **mayonnaise** or **mayonnaise-type dressing** was added to your sandwich breads or rolls, how much was usually added?

- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

62e. How often was **margarine** (including low-fat) added to your sandwich bread or rolls?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

62f. How often was **butter** (including low-fat) added to your sandwich bread or rolls?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

62g. Each time **margarine** or **butter** was added to your sandwich breads or rolls, how much was usually added?

- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 2 teaspoons
- ☐ More than 2 teaspoons

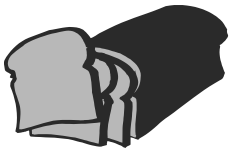
63. How often did you eat **breads** or **dinner rolls, NOT AS PART OF SANDWICHES**?

- ☐ NEVER (GO TO QUESTION 64)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

63a. Each time you ate **breads** or **dinner rolls, NOT AS PART OF SANDWICHES**, how much did you usually eat?

- ☐ 1 slice or 1 dinner roll
- ☐ 2 slices or 2 dinner rolls
- ☐ More than 2 slices or 2 dinner rolls



Question 64 appears on the next page

Over the <u>past 12 months</u>…

63b. How often were the breads or rolls you ate **white bread**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

63c. How often was **margarine** (including low-fat) added to your breads or rolls?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

63d. How often was **butter** (including low-fat) added to your breads or rolls?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

63e. Each time **margarine** or **butter** was added to your breads or rolls, how much was usually added?

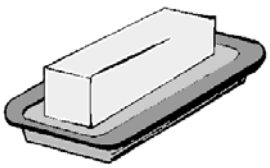
- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 2 teaspoons
- ☐ More than 2 teaspoons

63f. How often was **cream cheese** (including low-fat) added to your breads or rolls?

- ☐ Almost never or never (GO TO QUESTION 64)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

63g. Each time **cream cheese** was added to your breads or rolls, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 2 tablespoons
- ☐ More than 2 tablespoons



Question 64 appears in the next column

64. How often did you eat **jam, jelly,** or **honey** on bagels, muffins, bread, rolls, or crackers?

- ☐ NEVER (GO TO QUESTION 65)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

64a. Each time you ate **jam, jelly,** or **honey**, how much did you usually eat?

- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

65. How often did you eat **peanut butter** or **other nut butter**?

- ☐ NEVER (GO TO QUESTION 66)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

65a. Each time you ate **peanut butter** or **other nut butter**, how much did you usually eat?

- ☐ Less than 1 tablespoon
- ☐ 1 to 2 tablespoons
- ☐ More than 2 tablespoons

66. How often did you eat **roast beef** or **steak IN SANDWICHES**?

- ☐ NEVER (GO TO QUESTION 67)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

66a. Each time you ate **roast beef** or **steak IN SANDWICHES**, how much did you usually eat?

- ☐ Less than 1 slice or less than 2 ounces
- ☐ 1 to 2 slices or 2 to 4 ounces
- ☐ More than 2 slices or more than 4 ounces

Question 67 appears on the next page

Over the <u>past 12 months</u>…

67. How often did you eat **turkey** or **chicken COLD CUTS** (such as loaf, luncheon meat, turkey ham, turkey salami, or turkey pastrami)? (*We will ask about other turkey or chicken later.)*

☐ NEVER (GO TO QUESTION 68)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

67a. Each time you ate **turkey or chicken COLD CUTS**, how much did you usually eat?

☐ Less than 1 slice
☐ 1 to 3 slices
☐ More than 3 slices

68. How often did you eat **luncheon** or **deli-style ham**? *(We will ask about other ham later.)*

☐ NEVER (GO TO QUESTION 69)

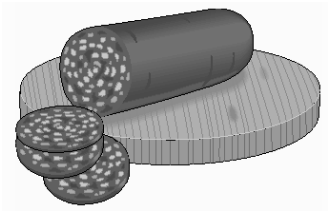
☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

68a. Each time you ate **luncheon** or **deli-style ham**, how much did you usually eat?

☐ Less than 1 slice
☐ 1 to 3 slices
☐ More than 3 slices

68b. How often was the luncheon or deli-style ham you ate **light**, **low-fat**, or **fat-free**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always



Question 69 appears in the next column

69. How often did you eat **other cold cuts** or **luncheon meats** (such as bologna, salami, corned beef, pastrami, or others, including low-fat)? *(Please do not include ham, turkey, or chicken cold cuts.)*

☐ NEVER (GO TO QUESTION 70)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

69a. Each time you ate **other cold cuts** or **luncheon meats**, how much did you usually eat?

☐ Less than 1 slice
☐ 1 to 3 slices
☐ More than 3 slices

69b. How often were the other cold cuts or luncheon meats you ate **light, low-fat**, or **fat-free cold cuts** or **luncheon meats**? *(Please do not include ham, turkey, or chicken cold cuts.)*

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

70. How often did you eat **canned tuna** (including in salads, sandwiches, or casseroles)?

☐ NEVER (GO TO QUESTION 71)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

70a. Each time you ate **canned tuna**, how much did you usually eat?

☐ Less than ¼ cup or less than 2 ounces
☐ ¼ to ½ cup or 2 to 3 ounces
☐ More than ½ cup or more than 3 ounces

70b. How often was the canned tuna you ate **water-packed tuna**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 71 appears on the next page

Over the <u>past 12 months</u>…

70c. How often was the canned tuna you ate **prepared with mayonnaise** or **other dressing** (including low-fat)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

71. How often did you eat **GROUND chicken** or **turkey**? *(We will ask about other chicken and turkey later.)*

- ☐ NEVER (GO TO QUESTION 72)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

71a. Each time you ate **GROUND chicken** or **turkey**, how much did you usually eat?

- ☐ Less than 2 ounces or less than ½ cup
- ☐ 2 to 4 ounces or ½ to 1 cup
- ☐ More than 4 ounces or more than 1 cup

72. How often did you eat **beef hamburgers** or **cheeseburgers**?

- ☐ NEVER (GO TO QUESTION 73)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

72a. Each time you ate **beef hamburgers** or **cheeseburgers**, how much did you usually eat?

- ☐ Less than 1 patty or less than 2 ounces
- ☐ 1 patty or 2 to 4 ounces
- ☐ More than 1 patty or more than 4 ounces

72b. How often were the beef hamburgers or cheeseburgers you ate made with **lean ground beef**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 73 appears in the next column

73. How often did you eat **ground beef in mixtures** (such as meatballs, casseroles, chili, or meatloaf)?

- ☐ NEVER (GO TO QUESTION 74)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

73a. Each time you ate **ground beef in mixtures**, how much did you usually eat?

- ☐ Less than 3 ounces or less than ½ cup
- ☐ 3 to 8 ounces or ½ to 1 cup
- ☐ More than 8 ounces or more than 1 cup

74. How often did you eat **hot dogs** or **frankfurters**? *(Please do not include sausages or vegetarian hot dogs.)*

- ☐ NEVER (GO TO QUESTION 75)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

74a. Each time you ate **hot dogs** or **frankfurters**, how many did you usually eat?

- ☐ Less than 1 hot dog
- ☐ 1 to 2 hot dogs
- ☐ More than 2 hot dogs

74b. How often were the hot dogs or frankfurters you ate **light** or **low-fat hot dogs**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 75 appears on the next page

Over the <u>past 12 months</u>…

75. How often did you eat beef mixtures such as **beef stew, beef pot pie, beef and noodles,** or **beef and vegetables**?

☐ NEVER (GO TO QUESTION 76)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

75a. Each time you ate **beef stew, beef pot pie, beef and noodles,** or **beef and vegetables,** how much did you usually eat?

☐ Less than 1 cup
☐ 1 to 2 cups
☐ More than 2 cups

76. How often did you eat **roast beef** or **pot roast**? *(Please do not include roast beef or pot roast in sandwiches.)*

☐ NEVER (GO TO QUESTION 77)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

76a. Each time you ate **roast beef** or **pot roast** (including in mixtures), how much did you usually eat?

☐ Less than 2 ounces
☐ 2 to 5 ounces
☐ More than 5 ounces

77. How often did you eat **steak** (beef)? (*Do not include steak in sandwiches*)

☐ NEVER (GO TO QUESTION 78)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

77a. Each time you ate **steak** (beef), how much did you usually eat?

☐ Less than 3 ounces
☐ 3 to 7 ounces
☐ More than 7 ounces

77b. How often was the steak you ate **lean steak**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

78. How often did you eat **pork** or **beef spareribs**?

☐ NEVER (GO TO QUESTION 79)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

78a. Each time you ate **pork** or **beef spareribs**, how much did you usually eat?

☐ Less than 4 ribs
☐ 4 to 12 ribs
☐ More than 12 ribs

79. How often did you eat **roast turkey, turkey cutlets,** or **turkey nuggets** (including in sandwiches)?

☐ NEVER (GO TO QUESTION 80)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

79a. Each time you ate **roast turkey, turkey cutlets,** or **turkey nuggets**, how much did you usually eat? *(Please note: 4 to 8 turkey nuggets = 3 ounces.)*

☐ Less than 2 ounces
☐ 2 to 4 ounces
☐ More than 4 ounces

80. How often did you eat **chicken** as part of **salads, sandwiches, casseroles, stews,** or **other mixtures**?

☐ NEVER (GO TO QUESTION 81)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

Question 78 appears in the next column

Question 81 appears on the next page

Over the <u>past 12 months</u>…

80a. Each time you ate **chicken** as part of **salads, sandwiches, casseroles, stews,** or **other mixtures,** how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 /2 cups
- ☐ More than 1 /2 cups

81. How often did you eat **baked, broiled, roasted, stewed,** or **fried chicken** (including nuggets)? *(Please do not include chicken in mixtures.)*

- ☐ NEVER (GO TO QUESTION 82)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

81a. Each time you ate **baked, broiled, roasted, stewed,** or **fried chicken** (including nuggets), how much did you usually eat?

- ☐ Less than 2 drumsticks or wings, less than 1 breast or thigh, or less than 4 nuggets
- ☐ 2 drumsticks or wings, 1 breast or thigh, or 4 to 8 nuggets
- ☐ More than 2 drumsticks or wings, more than 1 breast or thigh, or more than 8 nuggets

81b. How often was the chicken you ate **fried chicken** (including deep fried) or **chicken nuggets**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

81c. How often was the chicken you ate **WHITE meat**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

81d. How often did you eat chicken **WITH skin**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 82 appears in the next column

82. How often did you eat **baked ham** or **ham steak**?

- ☐ NEVER (GO TO QUESTION 83)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

82a. Each time you ate **baked ham** or **ham steak**, how much did you usually eat?

- ☐ Less than 1 ounce
- ☐ 1 to 3 ounces
- ☐ More than 3 ounces

83. How often did you eat **pork** (including chops, roasts, and in mixed dishes)? *(Please do not include ham, ham steak, or sausage.)*

- ☐ NEVER (GO TO QUESTION 84)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

83a. Each time you ate **pork**, how much did you usually eat?

- ☐ Less than 2 ounces or less than 1 chop
- ☐ 2 to 5 ounces or 1 chop
- ☐ More than 5 ounces or more than 1 chop

84. How often did you eat **gravy** on meat, chicken, potatoes, rice, etc.?

- ☐ NEVER (GO TO QUESTION 85)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

84a. Each time you ate **gravy** on meat, chicken, potatoes, rice, etc., how much did you usually eat?

- ☐ Less than ⅛ cup
- ☐ ⅛ to ½ cup
- ☐ More than ½ cup

Question 85 appears on the next page

Over the <u>past 12 months</u>…

85. How often did you eat **liver** (all kinds) or **liverwurst**?

☐ NEVER (GO TO QUESTION 86)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

85a. Each time you ate **liver** or **liverwurst**, how much did you usually eat?

☐ Less than 1 ounce
☐ 1 to 4 ounces
☐ More than 4 ounces

86. How often did you eat **bacon** (including low-fat)?

☐ NEVER (GO TO QUESTION 87)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

86a. Each time you ate **bacon**, how much did you usually eat?

☐ Fewer than 2 slices
☐ 2 to 3 slices
☐ More than 3 slices

86b. How often was the bacon you ate **light, low-fat,** or **lean bacon**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

87. How often did you eat **sausage** (including low-fat)?

☐ NEVER (GO TO QUESTION 88)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

87a. Each time you ate **sausage**, how much did you usually eat?

☐ Less than 1 patty or 2 links
☐ 1 to 3 patties or 2 to 5 links
☐ More than 3 patties or 5 links

87b. How often was the sausage you ate **light, low-fat**, or **lean sausage**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

88. How often did you eat **fish sticks** or **fried fish** (including fried seafood or shellfish)?

☐ NEVER (GO TO QUESTION 89)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

88a. Each time you ate **fish sticks** or **fried fish**, how much did you usually eat?

☐ Less than 2 ounces or less than 1 fillet
☐ 2 to 7 ounces or 1 fillet
☐ More than 7 ounces or more than 1 fillet

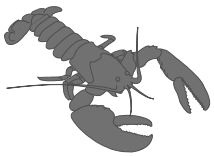
89. How often did you eat **fish** or **seafood that was NOT FRIED** (including shellfish)?

☐ NEVER (GO TO INTRODUCTION TO QUESTION 90)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

89a. Each time you ate eat **fish** or **seafood that was NOT FRIED**, how much did you usually eat?

☐ Less than 2 ounces or less than 1 fillet
☐ 2 to 5 ounces or 1 fillet
☐ More than 5 ounces or more than 1 fillet



Question 88 appears in the next column

Introduction to Question 90 appears on the next page

Over the <u>past 12 months</u>…

Now think about all the meat, poultry, and fish you ate in the <u>past 12 months</u> and how they were prepared.

90. How often was **oil, butter, margarine**, **or other fat** used to **FRY, SAUTE, BASTE, OR MARINATE** any meat, poultry, or fish you ate? *(Please do not include deep frying.)*

- ☐ NEVER (GO TO QUESTION 91)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

90a. Which of the following **fats** were regularly used to prepare your meat, poultry, or fish? ***(Mark all that apply.)***

- ☐ Margarine (including low-fat)
- ☐ Butter (including low-fat)
- ☐ Lard, fatback, or bacon fat
- ☐ Olive oil
- ☐ Corn oil
- ☐ Canola or rapeseed oil
- ☐ Oil spray, such as Pam or others
- ☐ Other kinds of oils
- ☐ None of the above

91. How often did you eat **tofu, soy burgers,** or **soy meat-substitutes**?

- ☐ NEVER (GO TO QUESTION 92)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

91a. Each time you ate **tofu, soy burgers,** or **soy meat-substitutes**, how much did you usually eat?

- ☐ Less than ¼ cup or less than 2 ounces
- ☐ ¼ to ½ cup or 2 to 4 ounces
- ☐ More than ½ cup or more than 4 ounces

Question 92 appears in the next column

92. Over the <u>past 12 months</u>, did you eat **soups**?

- ☐ NO (GO TO QUESTION 93)
- ☐ YES

92a. How often did you eat **soup DURING THE WINTER**?

- ☐ NEVER

- ☐ 1–6 times per winter
- ☐ 7–11 times per winter
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

92b. How often did you eat **soup DURING THE REST OF THE YEAR**?

- ☐ NEVER

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

92c. Each time you ate **soup**, how much did you usually eat?

- ☐ Less than 1 cup
- ☐ 1 to 2 cups
- ☐ More than 2 cups

92d. How often were the soups you ate **bean soups**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

92e. How often were the soups you ate **cream soups** (including chowders)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 93 appears on the next page

Over the <u>past 12 months</u>…

92f. How often were the soups you ate **tomato** or **vegetable soups**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

92g. How often were the soups you ate **broth soups** (including chicken) **with** or **without noodles** or **rice**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

93. How often did you eat **pizza**?

- ☐ NEVER (GO TO QUESTION 94)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

93a. Each time you ate **pizza**, how much did you usually eat?

- ☐ Less than 1 slice or less than 1 mini pizza
- ☐ 1 to 3 slices or 1 mini pizza
- ☐ More than 3 slices or more than 1 mini pizza

93b. How often did you eat pizza with **pepperoni, sausage**, or **other meat**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

94. How often did you eat **crackers**?

- ☐ NEVER (GO TO QUESTION 95)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 95 appears in the next column

94a. Each time you ate **crackers**, how many did you usually eat?

- ☐ Fewer than 4 crackers
- ☐ 4 to 10 crackers
- ☐ More than 10 crackers

95. How often did you eat **corn bread** or **corn muffins**?

- ☐ NEVER (GO TO QUESTION 96)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

95a. Each time you ate **corn bread** or **corn muffins,** how much did you usually eat?

- ☐ Less than 1 piece or muffin
- ☐ 1 to 2 pieces or muffins
- ☐ More than 2 pieces or muffins

96. How often did you eat **biscuits**?

- ☐ NEVER (GO TO QUESTION 97)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

96a. Each time you ate **biscuits**, how many did you usually eat?

- ☐ Fewer than 1 biscuit
- ☐ 1 to 2 biscuits
- ☐ More than 2 biscuits

97. How often did you eat **potato chips, tortilla chips,** or **corn chips** (including low-fat, fat-free, or low-salt)?

- ☐ NEVER (GO TO QUESTION 98)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 98 appears on the next page

Over the <u>past 12 months</u>…

97a. Each time you ate **potato chips, tortilla chips,** or **corn chips**, how much did you usually eat?

- ☐ Fewer than 10 chips or less than 1 cup
- ☐ 10 to 25 chips or 1 to 2 cups
- ☐ More than 25 chips or more than 2 cups

97b. How often were the chips you ate **Wow chips** or other **chips made with fat substitute** (Olean or Olestra)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

97c. How often were the chips you ate other **low-fat** or **fat-free chips**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

98. How often did you eat **popcorn** (including low-fat)?

- ☐ NEVER (GO TO QUESTION 99)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

98a. Each time you ate **popcorn**, how much did you usually eat?

- ☐ Less than 2 cups, popped
- ☐ 2 to 5 cups, popped
- ☐ More than 5 cups, popped

99. How often did you eat **pretzels**?

- ☐ NEVER (GO TO QUESTION 100)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 100 appears in the next column

99a. Each time you ate **pretzels**, how many did you usually eat?

- ☐ Fewer than 5 average twists
- ☐ 5 to 20 average twists
- ☐ More than 20 average twists

100. How often did you eat **peanuts, walnuts, seeds,** or **other nuts**?

- ☐ NEVER (GO TO QUESTION 101)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

100a. Each time you ate **peanuts, walnuts, seeds,** or **other nuts**, how much did you usually eat?

- ☐ Less than ¼ cup
- ☐ ¼ to ½ cup
- ☐ More than ½ cup

101. How often did you eat **energy, high-protein**, or **breakfast bars** such as **Power Bars, Balance, Clif,** or **others**?

- ☐ NEVER (GO TO QUESTION 102)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

101a. Each time you ate **energy, high-protein**, or **breakfast bars**, how much did you usually eat?

- ☐ Less than 1 bar
- ☐ 1 bar
- ☐ More than 1 bar

102. How often did you eat **yogurt** (NOT including frozen yogurt)?

- ☐ NEVER (GO TO QUESTION 103)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 103 appears on the next page

Over the <u>past 12 months</u>…

102a. Each time you ate **yogur**t, how much did you usually eat?

- ☐ Less than ½ cup or less than 1 container
- ☐ ½ to 1 cup or 1 container
- ☐ More than 1 cup or more than 1 container

103. How often did you eat **cottage cheese** (including low-fat)?

- ☐ NEVER (GO TO QUESTION 104)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

103a. Each time you ate **cottage cheese**, how much did you usually eat?

- ☐ Less than ¼ cup
- ☐ ¼ to 1 cup
- ☐ More than 1 cup

104. How often did you eat **cheese** (including low-fat; including on cheeseburgers or in sandwiches or subs)?

- ☐ NEVER (GO TO QUESTION 105)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

104a. Each time you ate **cheese**, how much did you usually eat?

- ☐ Less than ½ ounce or less than 1 slice
- ☐ ½ to 1 /2 ounces or 1 slice
- ☐ More than 1 /2 ounces or more than 1 slice

104b. How often was the cheese you ate **light** or **low-fat cheese**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

104c. How often was the cheese you ate **fat-free cheese**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

105. How often did you eat **frozen yogurt, sorbet,** or **ices** (including low-fat or fat-free)?

- ☐ NEVER (GO TO QUESTION 106)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

105a. Each time you ate **frozen yogurt, sorbet,** or **ices**, how much did you usually eat?

- ☐ Less than ½ cup or less than 1 scoop
- ☐ ½ to 1 cup or 1 to 2 scoops
- ☐ More than 1 cup or more than 2 scoops

106. How often did you eat **ice cream, ice cream bars,** or **sherbet** (including low-fat or fat-free)?

- ☐ NEVER (GO TO QUESTION 107)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

106a. Each time you ate **ice cream, ice cream bars,** or **sherbet**, how much did you usually eat?

- ☐ Less than ½ cup or less than 1 scoop
- ☐ ½ to 1 /2 cups or 1 to 2 scoops
- ☐ More than 1 /2 cups or more than 2 scoops

106b. How often was the ice cream you ate **light, low-fat,** or **fat-free ice cream** or **sherbet**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 105 appears in the next column

Question 107 appears on the next page

Over the <u>past 12 months</u>…

107. How often did you eat **cake** (including low-fat or fat-free)?

☐ NEVER (GO TO QUESTION 108)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

107a. Each time you ate **cake**, how much did you usually eat?

☐ Less than 1 medium piece
☐ 1 medium piece
☐ More than 1 medium piece

107b. How often was the cake you ate **light, low-fat,** or **fat-free cake**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

108. How often did you eat **cookies** or **brownies** (including low-fat or fat-free)?

☐ NEVER (GO TO QUESTION 109)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

108a. Each time you ate **cookies** or **brownies**, how much did you usually eat?

☐ Less than 2 cookies or 1 small brownie
☐ 2 to 4 cookies or 1 medium brownie
☐ More than 4 cookies or 1 large brownie

108b. How often were the cookies or brownies you ate **light, low-fat,** or **fat-free cookies** or **brownies**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 109 appears in the next column

109. How often did you eat **doughnuts, sweet rolls, Danish,** or **pop-tarts**?

☐ NEVER (GO TO QUESTION 110)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

109a. Each time you ate **doughnuts, sweet rolls, Danish,** or **pop-tarts**, how much did you usually eat?

☐ Less than 1 piece
☐ 1 to 2 pieces
☐ More than 2 pieces

110. How often did you eat **sweet muffins** or **dessert breads** (including low-fat or fat-free)?

☐ NEVER (GO TO QUESTION 111)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

110a. Each time you ate **sweet muffins or dessert breads**, how much did you usually eat?

☐ Less than 1 medium piece
☐ 1 medium piece
☐ More than 1 medium piece

110b. How often were the sweet muffins or dessert breads you ate **light, low-fat,** or **fat-free sweet muffins** or **dessert breads**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

111. How often did you eat **fruit crisp, cobbler,** or **strudel**?

☐ NEVER (GO TO QUESTION 112)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

Question 112 appears on the next page

Over the <u>past 12 months</u>…

111a. Each time you ate **fruit crisp, cobbler,** or **strudel,** how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

112. How often did you eat **pie**?

- ☐ NEVER (GO TO QUESTION 113)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

112a. Each time you ate **pie**, how much did you usually eat?

- ☐ Less than ⅛ of a pie
- ☐ About ⅛ of a pie
- ☐ More than ⅛ of a pie

The next four questions ask about the kinds of pie you ate. Please read all four questions before answering.

112b. How often were the pies you ate **fruit pie** (such as apple, blueberry, others)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

112c. How often were the pies you ate **cream, pudding, custard,** or **meringue pie**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

112d. How often were the pies you ate **pumpkin** or **sweet potato pie**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 113 appears in the next column

112e. How often were the pies you ate **pecan pie**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

113. How often did you eat **chocolate candy**?

- ☐ NEVER (GO TO QUESTION 114)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

113a. Each time you ate **chocolate candy**, how much did you usually eat?

- ☐ Less than 1 average bar or less than 1 ounce
- ☐ 1 average bar or 1 to 2 ounces
- ☐ More than 1 average bar or more than 2 ounces

114. How often did you eat **other candy**?

- ☐ NEVER (GO TO QUESTION 115)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

114a. Each time you ate **other candy**, how much did you usually eat?

- ☐ Fewer than 2 pieces
- ☐ 2 to 9 pieces
- ☐ More than 9 pieces

115. How often did you eat **eggs, egg whites,** or **egg substitutes** (NOT counting eggs in baked goods and desserts)? *(Please include eggs in salads, quiche, and soufflés.)*

- ☐ NEVER (GO TO QUESTION 116)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 116 appears on the next page

Over the <u>past 12 months</u>…

115a. Each time you ate **eggs**, how many did you usually eat?

- ☐ 1 egg
- ☐ 2 eggs
- ☐ 3 or more eggs

115b. How often were the eggs you ate **egg substitutes**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

115c. How often were the eggs you ate **egg whites only**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

115d. How often were the eggs you ate **regular whole eggs**?

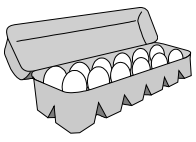
- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

115e. How often were the eggs you ate **cooked in oil, butter,** or **margarine**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

115f. How often were the eggs you ate part of **egg salad**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always



Question 116 appears in the next column

116. How many cups of **coffee**, caffeinated or decaffeinated, did you drink?

- ☐ NEVER (GO TO QUESTION 117)
- ☐ Less than 1 cup per month
- ☐ 1–3 cups per month
- ☐ 1 cup per week
- ☐ 2–4 cups per week
- ☐ 5–6 cups per week
- ☐ 1 cup per day
- ☐ 2–3 cups per day
- ☐ 4–5 cups per day
- ☐ 6 or more cups per day

116a. How often was the coffee you drank **decaffeinated**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

117. How many glasses of **ICED tea,** caffeinated or decaffeinated, did you drink?

- ☐ NEVER (GO TO QUESTION 118)
- ☐ Less than 1 cup per month
- ☐ 1–3 cups per month
- ☐ 1 cup per week
- ☐ 2–4 cups per week
- ☐ 5–6 cups per week
- ☐ 1 cup per day
- ☐ 2–3 cups per day
- ☐ 4–5 cups per day
- ☐ 6 or more cups per day

117a. How often was the iced tea you drank **decaffeinated** or **herbal tea**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

118. How many cups of **HOT tea**, caffeinated or decaffeinated, did you drink?

- ☐ NEVER (GO TO QUESTION 119)
- ☐ Less than 1 cup per month
- ☐ 1–3 cups per month
- ☐ 1 cup per week
- ☐ 2–4 cups per week
- ☐ 5–6 cups per week
- ☐ 1 cup per day
- ☐ 2–3 cups per day
- ☐ 4–5 cups per day
- ☐ 6 or more cups per day

118a. How often was the hot tea you drank **decaffeinated** or **herbal tea**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 119 appears on the next page

Over the <u>past 12 months</u>…

119. How often did you add **sugar** or **honey** to your coffee or tea?

☐ NEVER (GO TO QUESTION 120)

☐ Less than 1 cup per month
☐ 1–3 cups per month
☐ 1 cup per week
☐ 2–4 cups per week
☐ 5–6 cups per week
☐ 1 cup per day
☐ 2–3 cups per day
☐ 4–5 cups per day
☐ 6 or more cups per day

119a. Each time **sugar** or **honey** was added to your coffee or tea, how much was usually added?

☐ Less than 1 teaspoon
☐ 1 to 3 teaspoons
☐ More than 3 teaspoons

120. How often did you add **artificial sweetener** to your coffee or tea?

☐ NEVER (GO TO QUESTION 121)

☐ Less than 1 time per month
☐ 1–3 times per month
☐ 1 time per week
☐ 2–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

120a. What kind of **artificial sweetener** did you usually use?

☐ Equal or aspartame
☐ Sweet N Low or saccharin

121. How often was **non-dairy creamer** added to your coffee or tea?

☐ NEVER (GO TO QUESTION 122)

☐ Less than 1 time per month
☐ 1–3 times per month
☐ 1 time per week
☐ 2–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

121a. Each time **non-dairy creamer** was added to your coffee or tea, how much was usually used?

☐ Less than 1 teaspoon
☐ 1 to 3 teaspoons
☐ More than 3 teaspoons

Question 122 appears in the next column

121b. What kind of **non-dairy creamer** did you usually use?

☐ Regular powdered
☐ Low-fat or fat-free powdered
☐ Regular liquid
☐ Low-fat or fat-free liquid

122. How often was **cream** or **half and half** added to your coffee or tea?

☐ NEVER (GO TO QUESTION 123)

☐ Less than 1 time per month
☐ 1–3 times per month
☐ 1 time per week
☐ 2–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

122a. Each time **cream** or **half and half** was added to your coffee or tea, how much was usually added?

☐ Less than 1 tablespoon
☐ 1 to 2 tablespoons
☐ More than 2 tablespoons

123. How often was **milk** added to your coffee or tea?

☐ NEVER (GO TO QUESTION 124)

☐ Less than 1 time per month
☐ 1–3 times per month
☐ 1 time per week
☐ 2–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

123a. Each time **milk** was added to your coffee or tea, how much was usually added?

☐ Less than 1 tablespoon
☐ 1 to 3 tablespoons
☐ More than 3 tablespoons

123b. What kind of **milk** was usually added to your coffee or tea?

☐ Whole milk
☐ 2% milk
☐ 1% milk
☐ Skim, nonfat, or ½% milk
☐ Evaporated or condensed (canned) milk
☐ Soy milk
☐ Rice milk
☐ Other

Question 124 appears on the next page

Over the <u>past 12 months</u>…

124. How often was **sugar** or **honey** added to foods you ate? (Please do not include sugar in coffee, tea, other beverages, or baked goods.)

- ☐ NEVER (GO TO INTRODUCTION TO QUESTION 125)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

124a. Each time **sugar** or **honey** was added to foods you ate, how much was usually added?

- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

The following questions are about the kinds of margarine, mayonnaise, sour cream, cream cheese, and salad dressing that you eat. If possible, please check the labels of these foods to help you answer.

125. Over the <u>past 12 months</u>, did you eat **margarine**?

- ☐ NO (GO TO QUESTION 126)
- ☐ YES

125a. How often was the margarine you ate **regular-fat margarine** (stick or tub)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

125b. How often was the margarine you ate **light** or **low-fat margarine** (stick or tub)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 126 appears in the next column

125c. How often was the margarine you ate **fat-free margarine**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

126. Over the past 12 months, did you eat **butter**?

- ☐ NO (GO TO QUESTION 127)
- ☐ YES

126a. How often was the butter you ate **light** or **low-fat butter**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

127. Over the <u>past 12 months</u>, did you eat **mayonnaise** or **mayonnaise-type dressing**?

- ☐ NO (GO TO QUESTION 128)
- ☐ YES

127a. How often was the mayonnaise you ate **regular-fat mayonnaise**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

127b. How often was the mayonnaise you ate **light** or **low-fat mayonnaise**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 128 appears on the next page

Over the <u>past 12 months</u>…

127c. How often was the mayonnaise you ate **fat-free mayonnaise**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

128. Over the <u>past 12 months</u>, did you eat **sour cream**?

- ☐ NO (GO TO QUESTION 129)
- ☐ YES

128a. How often was the sour cream you ate **regular-fat sour cream**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

128b. How often was the sour cream you ate **light, low-fat,** or **fat-free sour cream**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

129. Over the <u>past 12 months</u>, did you eat **cream cheese**?

- ☐ NO (GO TO QUESTION 130)
- ☐ YES

129a. How often was the cream cheese you ate **regular-fat cream cheese**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

129b. How often was the cream cheese you ate **light, low-fat,** or **fat-free cream cheese**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

130. Over the <u>past 12 months</u>, did you eat **salad dressing**?

- ☐ NO (GO TO INTRODUCTION TO QUESTION 131)
- ☐ YES

130a. How often was the salad dressing you ate **regular-fat salad dressing** (including oil and vinegar dressing)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

130b. How often was the salad dressing you ate **light** or **low-fat salad dressing**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

130c. How often was the salad dressing you ate **fat-free salad dressing**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

The following two questions ask you to summarize your usual intake of vegetables and fruits. Please do not include salads, potatoes, or juices.

131. Over the <u>past 12 months</u>, how many servings of **vegetables** (not including salad or potatoes) did you eat per week or per day?

- ☐ Less than 1 per week
- ☐ 1–2 per week
- ☐ 3–4 per week
- ☐ 5–6 per week
- ☐ 1 per day
- ☐ 2 per day
- ☐ 3 per day
- ☐ 4 per day
- ☐ 5 or more per day

Question 130 appears in the next column

Over the past 12 months…

132. Over the past 12 months, how many servings of **fruit** (not including juices) did you eat per week or per day?

- ☐ Less than 1 per week
- ☐ 1–2 per week
- ☐ 3–4 per week
- ☐ 5–6 per week
- ☐ 1 per day
- ☐ 2 per day
- ☐ 3 per day
- ☐ 4 per day
- ☐ 5 or more per day

133. Over the past month, which of the following foods did you eat **AT LEAST THREE TIMES?** ***(Mark all that apply.)***

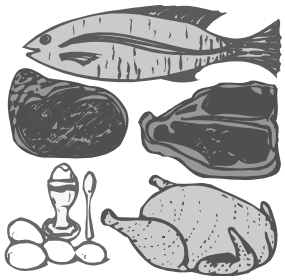
- ☐ Avocado, guacamole
- ☐ Cheesecake
- ☐ Chocolate, fudge, or butterscotch toppings or syrups
- ☐ Chow mein noodles
- ☐ Croissants
- ☐ Dried apricots
- ☐ Egg rolls
- ☐ Granola bars
- ☐ Hot peppers
- ☐ Jello, gelatin
- ☐ Milkshakes or ice-cream sodas
- ☐ Olives
- ☐ Oysters
- ☐ Pickles or pickled vegetables or fruit
- ☐ Plantains
- ☐ Pork neckbones, hock, head, feet
- ☐ Pudding or custard
- ☐ Veal, venison, lamb
- ☐ Whipped cream, regular
- ☐ Whipped cream, substitute
- ☐ NONE

134. For **ALL** of the past 12 months, have you followed any type of **vegetarian diet**?

- ☐ NO (GO TO INTRODUCTION TO QUESTION 135)
- ☐ YES

134a. Which of the following foods did you **TOTALLY EXCLUDE** from your diet? ***(Mark all that apply.)***

- ☐ Meat (beef, pork, lamb, etc.)
- ☐ Poultry (chicken, turkey, duck)
- ☐ Fish and seafood
- ☐ Eggs
- ☐ Dairy products (milk, cheese, etc.)



Introduction to Question 135 appears in the next column

The next questions are about your use of fiber supplements or vitamin pills.

135. Over the past 12 months, did you take any of the following types of **fiber or fiber supplements** on a regular basis (more than once per week for at least 6 of the last 12 months)? ***(Mark all that apply.)***

- ☐ NO, didn't take any fiber supplements on a regular basis (GO TO QUESTION 136)
- ☐ YES, psyllium products (such as Metamucil, Fiberall, Serutan, Perdiem, Correctol)
- ☐ YES, methylcellulose/cellulose products (such as Citrucel, Unifiber)
- ☐ YES, Fibercon
- ☐ YES, Bran (such as wheat bran, oat bran, or bran wafers)

136. Over the past 12 months, did you take any **multivitamins**, such as One-a-Day-, Theragran-, or Centrum-type multivitamins (as pills, liquids, or packets)?

- ☐ NO (GO TO INTRODUCTION TO QUESTION 138)
- ☐ YES

137. How often did you take **One-a-day-, Theragran-, or Centrum-type** multivitamins?

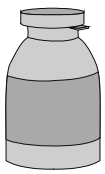
- ☐ Less than 1 day per month
- ☐ 1–3 days per month
- ☐ 1–3 days per week
- ☐ 4–6 days per week
- ☐ Every day

137a. Does your **multivitamin** usually contain **minerals** (such as iron, zinc, etc.)?

- ☐ NO
- ☐ YES
- ☐ Don't know

137b. For how many years have you taken **multivitamins**?

- ☐ Less than 1 year
- ☐ 1–4 years
- ☐ 5–9 years
- ☐ 10 or more years



Introduction to Question 138 appears on the next page

Over the <u>past 12 months</u>…

137c. Over the <u>past 12 months</u>, did you take any vitamins, minerals, or other herbal supplements other than your multivitamin?

☐ NO

Thank you *<u>very much</u>* for completing this questionnaire! Because we want to be able to use all the information you have provided, we would greatly appreciate it if you would please take a moment to review each page making sure that you:

- **Did not skip any pages and**
- **Crossed out the incorrect answer and circled the correct answer if you made any changes.**

☐ YES (GO TO INTRODUCTION TO QUESTION 138)

These last questions are about the vitamins, minerals, or herbal supplements you took that are <u>NOT</u> part of a One-a-day-, Theragran-, or Centrum-type of multivitamin.

Please include vitamins taken as part of an antioxidant supplement.

138. How often did you take **Beta-carotene** (**NOT** as part of a multivitamin in Question 137)?

☐ NEVER (GO TO QUESTION 139)

☐ Less than 1 day per month
☐ 1–3 days per month
☐ 1–3 days per week
☐ 4–6 days per week
☐ Every day

138a. When you took **Beta-carotene**, about how much did you take in one day?

☐ Less than 10,000 IU
☐ 10,000–14,999 IU
☐ 15,000–19,999 IU
☐ 20,000–24,999 IU
☐ 25,000 IU or more
☐ Don't know

138b. For how many years have you taken **Beta-carotene**?

☐ Less than 1 year
☐ 1–4 years
☐ 5–9 years
☐ 10 or more years

Question 139 appears in the next column

139. How often did you take **Vitamin A** (**NOT** as part of a multivitamin in Question 137)?

☐ NEVER (GO TO QUESTION 140)

☐ Less than 1 day per month
☐ 1–3 days per month
☐ 1–3 days per week
☐ 4–6 days per week
☐ Every day

139a. When you took **Vitamin A**, about how much did you take in one day?

☐ Less than 8,000 IU
☐ 8,000–9,999 IU
☐ 10,000–14,999 IU
☐ 15,000–24,999 IU
☐ 25,000 IU or more
☐ Don't know

139b. For how many years have you taken **Vitamin A**?

☐ Less than 1 year
☐ 1–4 years
☐ 5–9 years
☐ 10 or more years

140. How often did you take **Vitamin C** (**NOT** as part of a multivitamin in Question 137)?

☐ NEVER (GO TO QUESTION 141)

☐ Less than 1 day per month
☐ 1–3 days per month
☐ 1–3 days per week
☐ 4–6 days per week
☐ Every day

140a. When you took **Vitamin C**, about how much did you take in one day?

☐ Less than 500 mg
☐ 500–999 mg
☐ 1,000–1,499 mg
☐ 1,500–1,999 mg
☐ 2,000 mg or more
☐ Don't know

140b. For how many years have you taken **Vitamin C**?

☐ Less than 1 year
☐ 1–4 years
☐ 5–9 years
☐ 10 or more years

Question 141 appears on the next page

Over the <u>past 12 months</u>…

141. How often did you take **Vitamin E** (**NOT** as part of a multivitamin in Question 137)?

- ☐ NEVER (GO TO QUESTION 142)
- ☐ Less than 1 day per month
- ☐ 1–3 days per month
- ☐ 1–3 days per week
- ☐ 4–6 days per week
- ☐ Every day

141a. When you took **Vitamin E**, about how much did you take in one day?

- ☐ Less than 400 IU
- ☐ 400–799 IU
- ☐ 800–999 IU
- ☐ 1,000 IU or more
- ☐ Don't know

141b. For how many years have you taken **Vitamin E**?

- ☐ Less than 1 year
- ☐ 1–4 years
- ☐ 5–9 years
- ☐ 10 or more years

142. How often did you take **Calcium** or **Calcium-containing antacids** (**NOT** as part of a multivitamin in Question 137)?

- ☐ NEVER (GO TO QUESTION 143)
- ☐ Less than 1 day per month
- ☐ 1–3 days per month
- ☐ 1–3 days per week
- ☐ 4–6 days per week
- ☐ Every day

142a. When you took **Calcium** or **Calcium-containing antacids**, about how much elemental calcium did you take in one day? *(If possible, please check the label for elemental calcium.)*

- ☐ Less than 500 mg
- ☐ 500–599 mg
- ☐ 600–999 mg
- ☐ 1,000 mg or more
- ☐ Don't know

Question 143 appears in the next column

142b. For how many years have you taken **Calcium** or **Calcium-containing antacids**?

- ☐ Less than 1 year
- ☐ 1–4 years
- ☐ 5–9 years
- ☐ 10 or more years

The last two questions ask you about other supplements you took <u>more than once per week</u>.

143. Please mark any of the following **single supplements** you took <u>more than once per week</u> (**NOT** as part of a multivitamin in Question 137):

- ☐ B-6
- ☐ B-complex
- ☐ Brewer's yeast
- ☐ Cod liver oil
- ☐ Coenzyme Q
- ☐ Fish oil (Omega-3 fatty acids)
- ☐ Folic acid/folate
- ☐ Glucosamine
- ☐ Hydroxytryptophan (HTP)
- ☐ Iron
- ☐ Niacin
- ☐ Selenium
- ☐ Zinc

144. Please mark any of the following **herbal** or **botanical supplements** you took <u>more than once per week</u>.

- ☐ Aloe Vera
- ☐ Astragalus
- ☐ Bilberry
- ☐ Cascara sagrada
- ☐ Cat's claw
- ☐ Cayenne
- ☐ Cranberry
- ☐ Dong Kuai (Tangkwei)
- ☐ Echinacea
- ☐ Evening primrose oil
- ☐ Feverfew
- ☐ Garlic
- ☐ Ginger
- ☐ Ginkgo biloba
- ☐ Ginseng (American or Asian)
- ☐ Goldenseal
- ☐ Grapeseed extract
- ☐ Kava, kava
- ☐ Milk thistle
- ☐ Saw palmetto
- ☐ Siberian ginseng
- ☐ St. John's wort
- ☐ Valerian
- ☐ Other

Thank you *<u>very much</u>*** **for completing this questionnaire! Because we want to be able to use all the information you have provided, we would greatly appreciate it if you would please take a moment to review each page making sure that you:**

- **Did not skip any pages and**
- **Crossed out the incorrect answer and circled the correct answer if you made any changes.**

Study Objectives

The Lombardi Cancer Center at Georgetown University Medical Center, in collaboration with the Washington Hospital Center, is conducting a study on prostate cancer. The main goal of the project is to determine susceptibility to prostate cancer by evaluating a person's ability to repair DNA damage. For this purpose, the researchers are collecting small samples of blood, saliva, nail clipping and urine as well as information about family history, diet, exercise, drinking and smoking habits. These specimen and the collected information will be available to qualified medical researchers for studies examining biological factors linked to prostate cancer susceptibility.

Despite its morbidity and mortality, very little is known about the causes of prostate cancer. Clinical observations suggest that certain biological factors put individuals at increased risk for this disease. The ability to identify such risk factors will have a major impact on cancer prevention and treatment.

We are presently recruiting healthy men and prostate cancer patients to be participants in the study. The purpose is to compare a group of cancer-free subjects to prostate cancer patients in an effort to determine genetic susceptibility to the disease. You can advance prostate cancer research by joining the study. Our professional staff will make sure to accommodate your schedule and needs to ensure that this is a pleasant experience for you. In addition, you will be notified when the results of the study become available.

Research

Cancer research gives hope. Doctors and researchers at hospitals and medical centers all across the country are learning more about what causes prostate and are exploring ways to prevent it. They are also looking for better ways to detect, diagnose, and treat this disease.

When cancer is found and treated early, the chances for survival are better. The data collected in this study is analyzed for susceptibility in DNA repair and will be available to qualified researchers as a resource for discovery of prostate cancer biomarkers. These biomarkers may be able to identify susceptible subpopulations where cancer prevention, screening, and treatment methods may be focused. They will also help scientists and doctors develop advanced prevention methods leading to decreased occurrence of this disease.

What is Prostate Cancer?

In the United States prostate cancer is the most commonly diagnosed non-skin cancer among men and it is the second most common cause of cancer deaths. In recent years, prostate cancer has become a worldwide public health concern and disease incidence is increasing in all populations. For this reason it is essential that all risk factors possibly contributing to this disease are studied.

Cancer is a group of many different diseases that all arise in cells, the body's basic unit of life. The body is made up of many types of cells. Normally, cells grow and divide to produce more cells only when the body needs them. This orderly process helps keep the body healthy. Sometimes cells keep dividing when new cells are not needed. These cells may form a mass of extra tissue called a growth or tumor. Tumors can be benign or malignant.

How to Become Involved

You may become involved in this study if you:

- are living in the greater Washington DC area including Maryland and Virginia
- have no prior cancer history

OR

- are a prostate cancer patient
- are over the age of 18

Upon contact, we will inform you about the study and verify your eligibility to participate. We will collect information about your alcohol and tobacco history, occupational history, family history, diet and exercise. You will be asked to donate a small sample of blood, urine, saliva, nail clipping and the left over tumor tissue that may have been removed if you are a cancer patient. Contact us at any time if you need more information or decide to participate. You can enter the study right now as you are waiting in the clinic by calling the number below or by notifying clinic staff.

Principal Investigator: Radoslav Goldman, Ph.D.
Study Coordinator: Alexandra Schopf
Prostate Cancer Biomarker Resource
Lombardi Cancer Center
3800 Reservoir Road, NW
S-Level, Rm. 180
Washington, DC 20057-1465
Ph: (202) 687-0343
email: ajs57@georgetown.edu

MedStar Research Institute

Washington Hospital Center

Lombardi Cancer Center

Prostate Cancer Biomarker Resource Study

Bioinformatics Advance Access published September 13, 2005

BIOINFORMATICS

Analysis of Mass Spectral Serum Profiles for Biomarker Selection

Habtom W. Ressom[*1], Rency S. Varghese[1], Mohamed Abdel-Hamid[2], Sohair Abdel-Latif Eissa[3], Daniel Saha[1], Lenka Goldman[1], Emanuel F. Petricoin[3], Thomas P. Conrads[4], Timothy D. Veenstra[4], Christopher A. Loffredo[1] and Radoslav Goldman[1]

[1]Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA
[2]Viral Hepatitis Research Laboratory, NHTMRI, Cairo, Egypt
[3]National Cancer Institute, Cairo, Egypt
[4]Clinical Proteomics Program, NCI/FDA, Center for Biologics Evaluation, FDA, USA
[5]SAIC-Frederick and Biomedical Proteomics Program, NCI, Frederick, MD, USA

ABSTRACT

Motivation: Mass spectrometric profiles of peptides and proteins obtained by current technologies are characterized by complex spectra, high dimensionality, and substantial noise. These characteristics generate challenges in discovery of proteins and protein-profiles that distinguish disease states, e.g. cancer patients from healthy individuals. We present low-level methods for processing of mass spectral data and a machine learning method that combines support vector machines with particle swarm optimization for biomarker selection.

Results: The proposed method identified mass points that achieved high prediction accuracy in distinguishing liver cancer patients from healthy individuals in SELDI-QqTOF profiles of serum.

Availability: MATLAB scripts to implement the methods described in this paper are available from HWR's lab website at http://lombardi.georgetown.edu/labpage

Contact: hwr@georgetown.edu

1 INTRODUCTION

Mass spectrometric profiling of serum was optimized for high-throughput comparison of complex samples that allows discovery of biomarkers of diseases such as cancer (Petricoin *et al.*, 2002a). Independent analysis of the results pointed out the importance of avoiding bias and the need for independent validation of results (Baggerly *et al*. 2004; Diamandis, 2004; Ransohoff, 2005). Improved study design and technology in second-generation studies continue to indentify biomarker-candidates for variety of cancers including hepatocellular carcinoma (Zhang *et al*., 2004; Conrads *et al*., 2004, Paradis, 2005). This paper adds signal processing and biomarker selection methods to a growing number of improved tools for mass spectrometric identification of biomarkers in serum.

Data preprocessing such as smoothing, baseline correction, normalization, peak detection and peak alignment improve the performance of mass spectrometric data analysis methods for biomarker discovery (Sauve and Speed, 2004; Malyarenko *et al.*, 2005). The reason for this includes the substantial amount of noise and systematic variations between spectra caused by sample degradation over time, ionization suppression, and other parameters reviewed previously (Ransohoff, 2005; Semmes, 2005). Sorace and Zhan (2003) have reported the potential for non-biologic experimental bias in their assessment of ovarian cancer serum surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) profiling due to the presence of very low mass classifiers, which the authors concluded could not possibly come from biology. This contention was disputed since many low molecular weight molecules detected by mass spectrometry could be metabolites or entities such as lysophosphatidic acid, a potential biomarker for ovarian cancer detection with a mass of 430 Da (Petricoin *et al*., 2004).

Mass spectra represent a complex signal consisting of electronic noise, chemical noise due to contaminants and matrix, and protein and metabolic signatures (Petricoin *et al.*, 2002b). They also have a varying baseline caused, besides others, by matrix-associated chemical noise or by ion overload**.** The latter refers to the high excess of ions derived from the matrix that can overload the detector (Malyarenko *et al.*, 2005). This elevates the baseline from its ideal zero horizontal line.

Previous quality-control experiments have suggested several measurement properties of current mass spectrometry technologies that must be accounted for in the analysis (Yasui *et al.*, 2003). These properties include high dimensionality of the spectra, high coefficients of variation, and mass shift (measurement error) Thus, it is important to apply low-level analyses that enable the recognition of spectral quality prior to using the spectra for biomarker discovery and disease classification. The low-level corrections are typically available in every software for operation of a mass spectrometer. The use of spectral comparisons for biomarker identification requires, however, optimization of these methods and a completely transparent data manipulation. Several groups proposed recently improved tools for signal processing for biomarker discovery as summarized briefly below.

By smoothing the raw spectra, we can reduce the effect of some mass-per-charge (m/z) values that appear as peaks but may not be or are very hard to verify by independent experiments. Many smoothing algorithms are available to denoise raw signals including the well-known Savitzky-Golay filter that removes additive

* To whom correspondence should be addressed.

white noise (Pusztai *et al.*, 2004) and wavelets (Coombes *et al.*, 2004).

Baseline correction is important for minimization of background noise; drifting baseline introduces serious distortion of ion intensities without adequate correction. Several methods have been proposed for baseline subtraction. For example, Fung and Enderwick (2002) employed a varying-width segemented convex hull algorithm to subtract the baseline. Baggerly *et al.* (2003) fitted a local median or local mean in a fixed window on the time scale. They also considered subtracting a "semimonotonic" baseline. Coombes *et al.* (2004) estimated baseline by fitting a monotone local minimum curve to smoothed spectra.

Normalization reduces variation in signal intensity between spectra. A commonly used normalization method for mass spectrometric data is rescaling each spectrum by its total ion current, i.e., the area under the curve (AUC) (Fung and Enderwick, 2002; Sauve and Speed, 2004). Other common choices for the rescaling coefficient include the spectrum median or mean. Alternatively, choosing the average AUC over all spectra as the rescaling coefficient can do a global normalization. A global optimization assumes that the sample intensities are all related by a constant factor. That means that the data distribution should not differ substantially from one spectrum to another.

Peak detection deals with the selection of m/z values which display a reasonable intensity compared to those that appear noise. Coombes *et al.* (2004) applied a simple peak finding (SPF) algorithm that provides the locations of potential peaks and their associated left-hand and right-hand bases. They estimated signal-to-noise ratio (S/N) using wavelets for improved peak detection. Also, they introduced a method for coalescing neighboring peaks.

Assuming appropriate low-level analysis methods are used for mass spectral data preprocessing, biomarker selection can be addressed using various computational methods. One of the commonly used approaches is to apply statistical analyses that recognize differentially expressed m/z values between cases and controls with multiple subjects. For example, one can apply a two-sample t-test method to compare the protein intensities at each m/z value in cases and controls. Zhu *et al.* (2003) proposed a statistical algorithm that can select a subset of *k* biomarkers from the marker list that could best discriminate between the groups in a training dataset via the best *k*-subset discriminant method with high sensitivity and specificity.

Machine learning methods have also been proposed for biomarker discovery. For example, Petricoin *et al.* (2002a) applied a combination of genetic algorithm (GA) and self-organizing clustering (GA-SOC) for variable selection. The GA-SOC, which is implemented in ProteomeQuest software, starts with hundreds of random choices of small sets of exact m/z values selected from the SELDI-TOF mass spectra. Each candidate subset contains 5 to 20 of the potential m/z values that define the spectra. The m/z values within the highest rated sets are reshuffled to form new subset candidates. The candidates are rated iteratively until the set that fully discriminates the preliminary set emerges.

Koopmann *et al.* (2004) applied successfully support vector machines (SVMs) in a modified form to proteomic profiling. Li *et al.* (2002) introduced unified maximum separability analysis (UMSA) algorithm, which incorporates data distribution information into structural risk minimization learning algorithm. UMSA is applied to identify a direction along which two classes of data are best separated. This direction is represented as a linear combination of the original variables. The weight assigned to each variable in this combination measures the contribution of the variable toward the separation of the two classes of data. They analyzed protein profiles of serum samples from patient with or without breast cancer. They reported that UMSA enabled the identification of three discriminatory biomarkers that achieved 93% sensitivity and 91% specificity in detecting breast cancer patients from the non-cancer controls.

In our previous work (Ressom *et al.*, 2005), we proposed a novel computational method known as PSO-SVM that combines SVMs and particle swarm optimization (PSO) for optimal selection of m/z values from high resolution SELDI-quadrupole-TOF (SELDI-QqTOF) spectra. First, we performed binning, normalization, baseline correction, and peak identification. Then, we refined the identified peak list based on S/N of peaks and their frequency of occurrence in multiple spectra. Finally, we used the PSO-SVM algorithm to select optimal m/z values associated to the refined peak list.

In this paper, we performed peak alignment by combining neighboring peaks within a spectrum and across spectra. This peak alignment method defines windows of m/z values that have variable width. The PSO-SVM algorithm is applied to select the optimal m/z windows. We ran the algorithm multiple times and selected 7 to 9 m/z windows based on their frequency of occurrence. An SVM classifier that employs these m/z windows as its inputs yielded up to 91% sensitivity and 92% specificity in distinguishing hepatocellular carcinoma (HCC) patients from matched controls.

2 METHODS

2.1 Mass Spectral Data

The incidence of HCC in the United States increases. HCC has been associated with hepatitis C (HCV) and B (HBV) viral infections. Very high rates of HCC incidence are observed in Egypt where an epidemic of viral infections presents a serious health problem. The management of the disease would benefit from identification of biomarkers related to this disease. Serum samples of HCC cases and controls were obtained from 2000 to 2002 in collaboration with the National Cancer Institute of Cairo University, Egypt. Controls were recruited among patients from the orthopedic fracture clinic at the Kasr El-Aini Hospital, Cairo, Egypt and were frequency-matched by gender, rural versus urban birthplace, and age to cancer cases (Ezzat *et al*., 2005). Blood samples were collected by trained phlebotomist each day around 10am and processed within a few hours according to a standard protocol. Aliquots of sera for mass spectrometric analysis were frozen at -80^{o}C immediately after collection till analysis; all measurements were performed on samples of second-time thawed serum.

411 sera (199 from HCC patients and 212 from matched healthy individuals) were analyzed by using SELDI-QqTOF, a hybrid quadrupole time-of-flight (QSTAR, Applied Biosystems) mass spectrometer interfaced with a weak cation exchange (WCX) protein array (Ciphergen Biosystems). The protein array consists of eight spots at which samples are presented to the ionization source of the instrument. We used one of the spots for a reference serum and the rest for sera from cases and controls. Note that the same serum was used as a reference on each array throughout the study. To control bias that maybe introduced by spot location, we ana-

lyzed 3 cases and 4 controls (or 3 controls and 4 cases) on each array. The spot locations were interchanged in consecutive arrays. For example, if a spot location was used for a case in one array, it was used for a control or a reference in the next array. The spot location of the reference serum was determined on a rolling basis, i.e., it was moved from 1 to 8 and back to 1 in each consecutive run.

The replicate spectra of the reference serum were used to assess technical variability. 61 reference spectra were available for this study. Each spectrum had 6107 intensity values (found using a binning procedure discussed in the next section). We transformed each intensity value by computing the base-two logarithm and found the mean log intensity value and standard deviation. The coefficient of variation of the log-transformed intensity values in the 61 reference spectra ranged between 6.4% and 22.4% with mean value of 15.5%.

2.2 Low-level Analysis

We applied low-level analysis methods to preprocess the raw high-resolution SELDI-QqTOF mass spectra. We began our analysis with outlier screening where we removed spectra whose data distribution substantially deviated from others. To reduce the noise and dimensionality of the raw spectra, we used a binning procedure that divides the m/z axis into intervals of desired length. The mean of the intensities within each interval was used as the protein expression variable in each bin. The low-frequency baseline of each spectrum was estimated by using multiple shifted windows of 200 bins. Spline approximation was used to regress the varying baseline. The regressed baseline was subtracted from the spectrum yielding a baseline corrected spectrum. Each spectrum was normalized by dividing it by its total ion current. In addition, the spectra were scaled to have an overall maximum intensity of 100. For peak detection, a bin is identified as a peak if the sign of the intensity's slope changes from positive to negative. Peaks with intensity below a pre-defined threshold-line were considered as noise and were discarded. To account for variation in the m/z location (drifts) in different spectra, two peaks were coalesced if they differed in location by at most 2 bins or at most 0.08% relative mass. This method was based on the ideas of Coombes *et al*. (2004) who used this method for SELDI-TOF spectra, where they combined peaks if they fall within 7 clock ticks and differ by at most 0.3% relative mass.

2.3 Biomarker Selection

The purpose of this analysis is to identify optimal m/z windows or candidate biomarkers from the preprocessed mass spectral data. While peak detection deals with the selection of mass points with reasonable intensity and S/N ratio, the aim of biomarker selection is to identify mass points that can be used to distinguish between cancer patients and healthy individuals.

The PSO-SVM algorithm described in (Ressom *et al*., 2005) is used to select optimal m/z windows. The algorithm builds SVM classifiers for each potential solution generated by PSO. The prediction capability of the resulting SVM classifier on a validation dataset is used as a performance function for the PSO algorithm. Since SVMs provide good generalization capability in classification tasks and can be designed in a computationally efficient manner, they are an ideal candidate for use as a performance function.

The preprocessed data are split into training and testing (independent) datasets. The training dataset is used to select m/z windows and build an SVM classifier. The validity of each classifier trained with the selected features is evaluated using the prediction accuracy of the SVM classifier in distinguishing cancer patients from non-cancer controls. SVM classifiers are built for various combinations of features until the performance of the SVM classifier converges or a pre-specified maximum iteration number is reached.

Estimates of prediction accuracy are calculated by using the *k*-fold cross-validation and bootstrapping methods. In *k*-fold cross-validation, we divide the training dataset into *k* subsets of (approximately) equal size. We train the SVM classifier *k* times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the prediction accuracy. In bootstrapping, instead of analyzing pre-specified subsets of the training dataset, we repeatedly select subsamples of the data. Each subsample is a random sample with replacement from the full training dataset.

The PSO-SVM algorithm is used to identify the optimal m/z windows from a list of *L* potential m/z windows. The algorithm creates *N* vectors (particles), each consisting of *n* m/z windows that are randomly selected from *L* m/z windows. The algorithm evaluates the performance of each particle in distinguishing cases and controls. This is carried out by building an SVM classifier for each particle and evaluating the performance of the classifier via the *k*-fold cross-validation or bootstrapping methods. The algorithm uses the most-fit particles to contribute to the next generation of *N* candidate particles. Thus, on the average, each successive population of candidate particles fits better than its predecessor. This process continues until the performance of the SVM classifier converges. The algorithm repeats the above steps multiple times and provides a list of selected m/z windows along with their frequency of occurrence. A frequency plot is used to estimate the optimal number of m/z windows. The frequency plot presents the number of occurrences versus the m/z windows sorted in the order of decreasing frequency. We considered as candidate biomarkers all m/z windows starting from the first until the frequency curve becomes flat (i.e. the change in frequency becomes low). These m/z windows are evaluated via testing dataset (i.e., independent dataset that was used neither for training nor for variable selection) to determine the generalization capability of the SVM classifier.

3 RESULTS

3.1 Low-Level Analysis

About 13% of the 411 SELDI-QqTOF spectra displayed substantial deviation from the data distribution and were excluded, leaving 357 (176 cases and 181 controls) spectra for further analysis. These outliers were singled out based on their deviation from the median ion current, median record count (number of mass points), and their alignment with pre-selected landmarks.

Each spectrum consisted of ~340,000 m/z values with the corresponding ion intensities. The dimension of these high-resolution spectra was reduced to 6107 m/z values via a binning procedure that divides the m/z axis into intervals of desired length over the mass range 1 to 11.5 kDa. A bin size of 400 parts per million (ppm) was found adequate as it is 10 times the routine mass accu-

racy of the SELDI-QqTOF with external calibration. The mean of the intensities within each interval was used as the protein expression variable in each bin. The baseline of each spectrum was estimated by using a shifting window size of 200 bins. The baseline (background value) at every window is estimated by taking the 10% quantile value. Spline approximation was used to regress the varying baseline. The regressed baseline was subtracted from the spectrum yielding a baseline corrected spectrum. Spline regression estimates different linear slopes for different ranges of the m/z values. Eilers and Marx (1996) applied the method for baseline correction of 2-D gel electrophoresis images.

Furthermore, each spectrum was normalized by dividing it by its total ion current. Fig. 1 depicts a SELDI-QqTOF mass serum spectrum of a healthy individual. On the horizontal axis are m/z values and on the vertical axis are intensity measurements that indicate the relative ion abundance. The top figure is the raw spectrum and the bottom figure depicts the spectrum after binning. As shown in the figures, the binning algorithm has removed the high frequency noise, thus smoothing the spectrum. Also, it improved the alignment of the spectra (not shown). Fig. 2 (top) depicts the regressed baseline of the spectrum. The baseline corrected and normalized spectrum is shown in Fig. 2 (bottom).

The 176 case and 181 control spectra were split into training and testing datasets. The training dataset consisted of 200 samples (100 cases and 100 controls). The testing dataset had 157 samples (76 cases and 81 controls). We used the training dataset for peak detection and peak alignment. The training spectra were rescaled so that the maximum intensity across all spectra is 100. A bin is identified as a peak if a change in the sign of the intensity's slope occurred. Note that peaks with intensity below a pre-defined threshold-line were considered as noise and were discarded. In this study, we used a threshold-line that decreases linearly from 1.5 % of the overall maximum intensity to 0.1% as m/z increases from 1 kDa to 11.5 kDa. Thus, intensity values below this threshold line were considered as noise. This threshold-line was selected by visual inspection for this dataset; the threshold-line would be at a higher intensity level if the spectra were not smoothed via binning. The threshold line decreases with an increasing m/z because of the observed decrease in the noise-level at higher m/z values.

To accommodate drifts introduced by the instrument, we aligned peaks by coalescing neighboring peaks within and across spectra into m/z windows. First, we selected peaks above a threshold-line that decreases linearly from 2.5% to 1%. Then, we combined these peaks if they differed in location by at most 2 bins or at most 0.08% relative mass. This step found 444 m/z windows in the training dataset. Following this, we considered peaks with intensities between the threshold-line that decreases from 2.5% to 1% and another threshold-line, which decreases from 1.5% to 0.1%. These peaks were added into previously identified m/z windows if they fell within 2 bins or at most 0.08% relative mass. Note that this step may increase the width of an m/z window if a peak is added from outside, otherwise the m/z window size remains unchanged except that the number of peaks in that window will increase. We retained m/z windows that consisted of peaks from at least five spectra and discarded the rest. This step resulted in 368 m/z windows that satisfied the criterion. Finally, we found the maximum intensity within the 368 m/z windows for each spectrum in the training dataset, yielding a 368 x 200 training data matrix. The testing spectra were binned, baseline corrected, and normalized in the same way as the training spectra. They were rescaled based on the parameters used to rescale the training spectra, so that the maximum intensity in the training dataset is 100. The 368 m/z windows defined by the training spectra were used to create a 368 x 157 testing data matrix. Fig. 3 shows the training spectra and the m/z windows found in the range from 7.6 to 7.8 kDa.

3.2 Biomarker Selection Using PSO-SVM

We used the PSO-SVM algorithm to select candidate biomarkers from the 368 peak-containing m/z windows. In this study, we arbitrarily targeted selection of five m/z windows. The algorithm began with 50 particles, where each particle consisted of 5 randomly selected m/z values from the 368 windows (i.e., n = 5, N =50, and L = 368). A linear SVM classifier was built for each particle via the training dataset. The prediction power of each particle (5 m/z windows) was evaluated by measuring the performance of the SVM classifier in distinguishing the two classes through the k-fold cross validation and bootstrapping methods. We used k=10 for this study. The most-fit particles contributed to the next generation of 50 candidate particles. This process continued until the performance of the SVM classifier converged or a pre-specified number of iterations was reached. The algorithm was repeated 350 times, 175 runs using the 10-fold cross validation method and 175 runs using the bootstrapping method. Fig. 4 depicts the percentage of occurrence of m/z windows selected by the PSO-SVM. Note that the m/z windows are sorted in decreasing order of frequency and only the first 60 m/z windows are shown in the figure. As shown in the figure, the change in frequency of occurrence became small after the first 7 m/z windows. Each of the first 7 m/z windows (7918.8-7922, 8140.4-8143.6, 7746.5, 7887.2-7893.5, 7830.6-7836.9, 7934.7-7937.8, and 7728.0-7731.1) was selected in at least 15% of the runs. These m/z windows yielded 91% sensitivity and 88% specificity in distinguishing HCC patients from healthy individuals in the testing dataset.

We observed that the population size has no significant effect on feature selection. This is evident from the frequency plot shown in Fig. 5 , where we ran the PSO-SVM algorithm 350 times for n=5 and N=100. The frequency plot for this experiment yielded 8 potential markers with 91% sensitivity and 90 % specificity. The first 7 markers are the same as those found in the previous experiment with n=5 and N=50. Fig. 6 depicts these 7 m/z windows along with the training spectra, mean spectrum for cases, and mean spectrum for controls.

For comparison, we applied a two-sample t-test method to identify differentially expressed m/z windows from the preprocessed spectra. The method selected 128 m/z windows out of 368 at significance level $p < 0.001$. An SVM trained with the 128 m/z windows yielded 90% sensitivity and 90% specificity in the independent dataset. The eight m/z windows selected by the PSO-SVM in the above experiment were also selected by the t-test method as a part of 128 m/z windows. Each of these m/z windows is differentially expressed with $p < 0.0001$ and has a fold change > 2 between controls and cases. However, there were 122 m/z windows with $p < 0.0001$ and 38 with fold change > 2. This demonstrates the power of the PSO-SVM algorithm in identifying a small set of relevant candidate biomarkers despite the presence of large number of statistically significant potential candidates.

To examine any potential bias that may be introduced by parameter choice, we ran the PSO-SVM algorithm 600 times for

various numbers of features (n=5 to 10) and particles (N=50 and 100). Fig. 7 depicts the frequency plot for this experiment. As shown in the figure, 7 out of the first 9 potential markers are the same as those found in the previous two experiments. These 9 m/z windows distinguished the HCC patients from healthy individuals in the testing dataset with 91% sensitivity and 92% specificity.

To study the effect of preprocessing, we performed biomarker selection using spectra that were binned and normalized, but not baseline corrected. 684 m/z windows were found from these spectra using our peak detection and alignment methods described before. The increase in the number of m/z windows is attributed to features that were not baseline corrected. The PSO-SVM algorithm was run 120 times for features ranging from n=5 to 10, with N = 50 and 100 particles. The corresponding frequency plot provided five m/z windows, of which four were the same as those found in the above three experiments. These 5 m/z windows yielded 85% sensitivity and 90% specificity. This shows that baseline correction has an impact in improving the prediction accuracy. Table 1 summarizes the improvement in classification performance with baseline correction for various features (all bins, all m/z windows, and the m/z windows selected by the PSO-SVM algorithm when n was varied between 5 and 10).

Table 1: Comparison of classification performance with and without baseline correction

Features	without baseline correction		with baseline correction	
	Sen.	Spe.	Sen.	Spe.
all m/z bins	91	85	90	90
all m/z windows	90	83	92	89
m/z windows selected by PSO-SVM	85	90	91	92

4 CONCLUSIONS AND FUTURE WORK

In this paper, we presented low-level analysis methods for mass spectral data preprocessing. A computational method that combines particle swarm optimization with support vector machines is applied for biomarker selection. We showed that the proposed approaches can select mass points from the complex mass spectra. For the SELDI-QqTOF data presented in this paper, 7 to 9 m/z windows were selected that yielded up to 91% sensitivity and 92% specificity in distinguishing liver cancer patients from healthy individuals in an independent dataset. Compared to our previous study for the same data, we observed that the use of m/z windows provides equal or better performance than precise m/z values or m/z bins. The m/z windows selected by the PSO-SVM algorithm consist of clearly detectable peaks, which are more likely to represent identifiable proteins, protein fragments or peptides. This is important for our ultimate goal of identifying proteins/peptides that distinguish cancer patients from healthy individuals. Once the proteins are identified, focus will be on validating the proteins through other sample-sets and analytical platforms such as the increasingly popular matrix-assisted laser-desorption/ionization time-of-flight (MADLI-TOF) mass spectrometry (Villanueva *et al.*, 2004; Koomen *et al.*, 2005).

We believe that the use of computational methods alone cannot provide a solution to the complex task of biomarker discovery from mass spectra involving thousands of proteins. In addition to advanced computational methods that are capable of extracting knowledge from complex and high dimensional data, this task requires careful study design, sample collection and preparation, improved mass spectrometry, well-designed low-level analyses, and inter-laboratory validation.

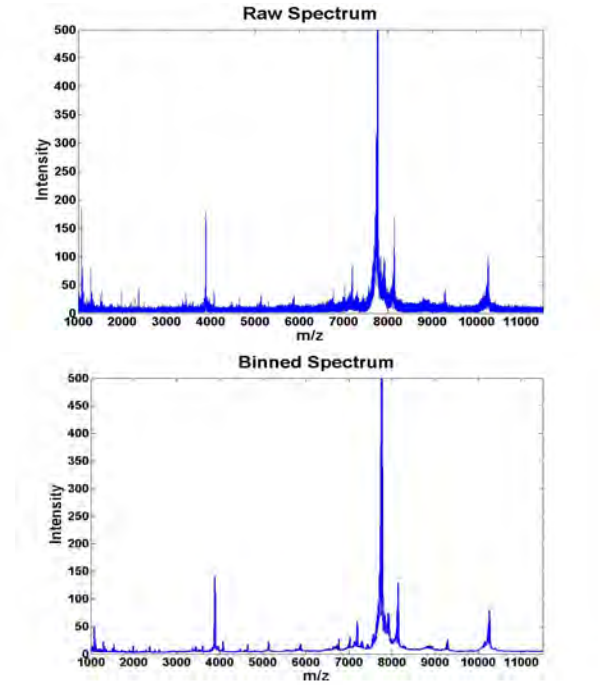


Fig. 1. SELDI-QqTOF mass spectrum in the range between 1-11.5 kDa: raw (top figure) and binned (bottom figure).

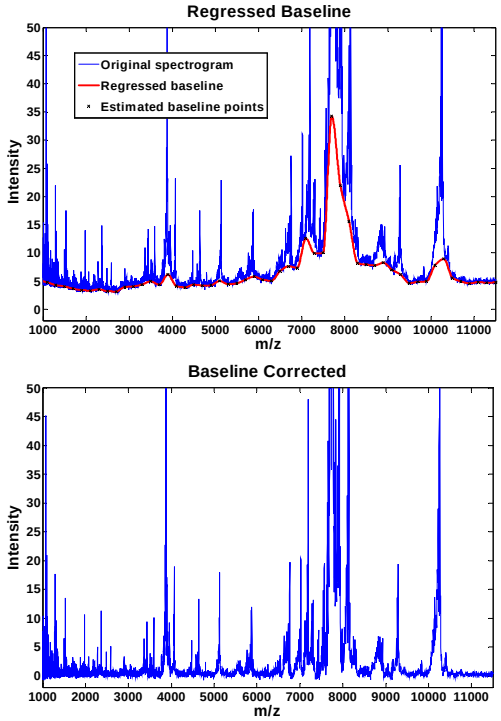


Fig. 2. SELDI-QqTOF normalized spectrum and regressed baseline (top figure) and baseline corrected (bottom figure). Note that the above spectra show intensity values in the range 0-50.

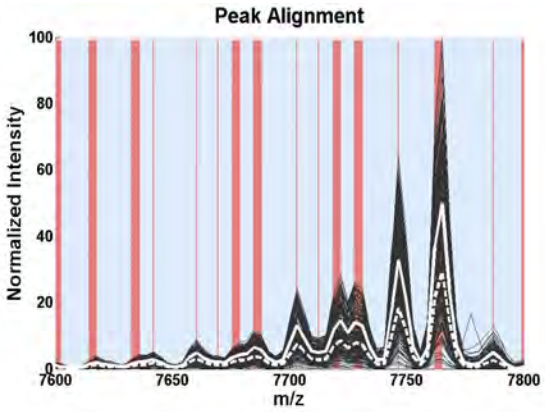


Fig. 3. Training spectra (black solid lines), windows in the m/z range from 7.6 to 7.8 kDa along with averaged control spectrum (white solid line) and averaged case spectrum (white broken line).

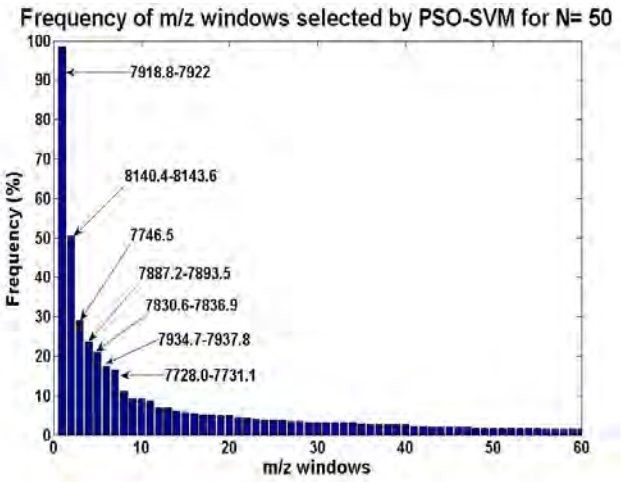


Fig. 4. Frequency of occurrence of m/z windows for N = 50 in 350 PSO-SVM runs sorted in decreasing order of frequency (only the first 60 m/z windows are shown).

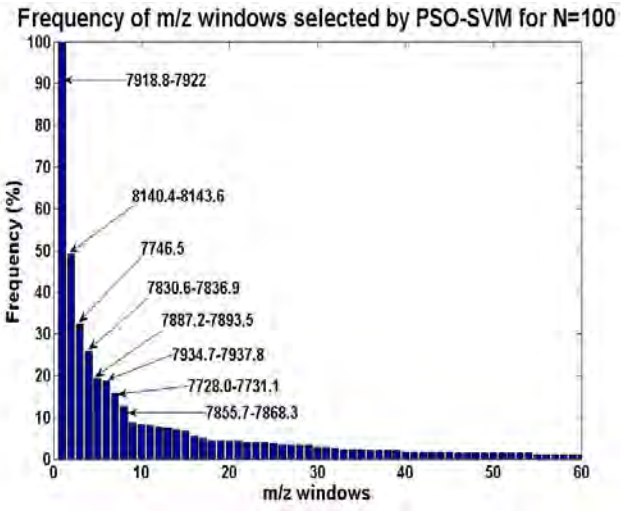


Fig. 5. Frequency of occurrence of m/z windows for N = 100 in 350 PSO-SVM runs sorted in decreasing order of frequency (only the first 60 m/z windows are shown).

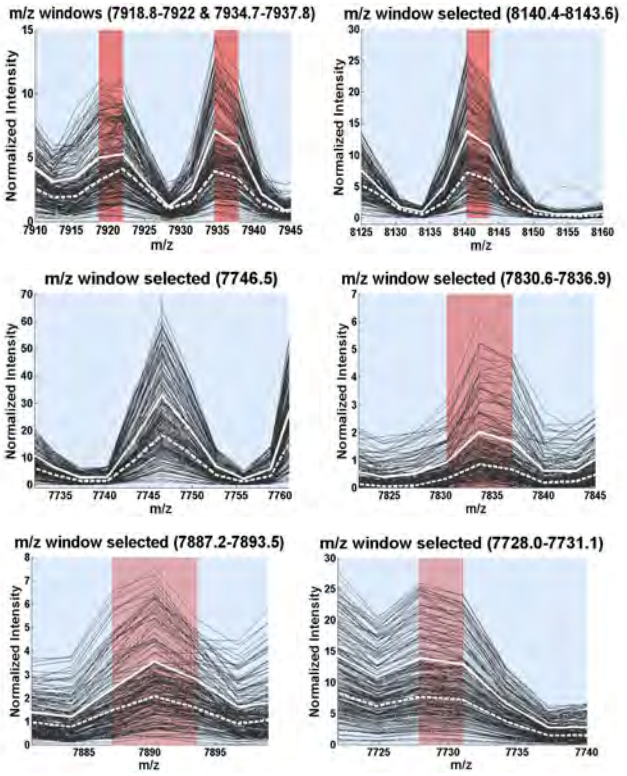


Fig. 6. m/z windows selected by PSO-SVM training spectra (black solid lines), averaged control spectrum (white solid line), and averaged case spectrum (white broken line).

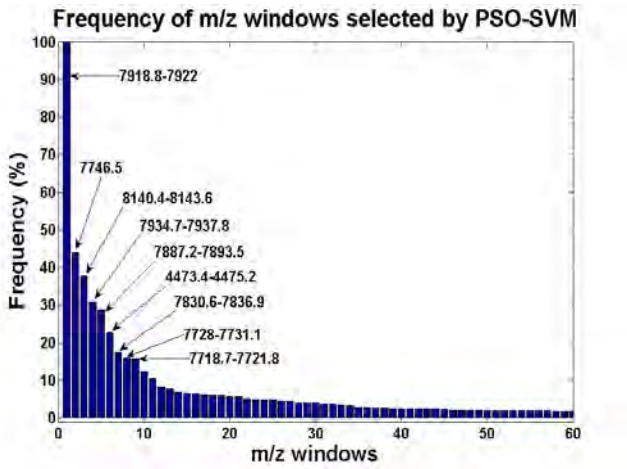


Fig. 7. Frequency of occurrence of m/z windows for n=5-10 and N=50 and 100 in 600 PSO-SVM runs sorted in decreasing order of frequency (only the first 60 m/z windows are shown).

ACKNOWLEDGEMENTS

This work was supported in part by U.S. Army Medical Research and Material Command, Prostate Cancer Research Program grant DAMD17-02-1-0057 and American Cancer Society grant CRTG-02-245-01-CCE awarded to RG.

We thank J. Jelinek for the programming of the binning algorithm. We thank Drs. F. Seillier-Moiseiwitsch and A. Wang for their helpful suggestions and discussions.

REFERENCES

Baggerly, K.A., Morris, J.S., and Coombes, K.R. (2004) Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinf.* 20(5):777-85.

Baggerly, K.A., Morris, J.S., Wang, J., Gold, D., Xiao, L.C., and Coombes, K.R. (2003) A comprehensive approach to the analysis of MALDI-TOF proteomics spectra from serum samples. *Proteomics*, **3** (9), 1667-1672.

Conrads, T.P., Fusaro, V.A., Ross, S., Johann, D., Rajapakse, V., Hitt, B.A., Steinberg, S.M., Kohn, E.C., Fishman, D.A., Whitely, G., Barrett, J.C., Liotta, L.A., Petricoin, E.F., and Veenstra. T.D. (2004) High-resolution serum proteomic features for ovarian cancer detection. *Endocr Relat Cancer* 11(2):163-178.

Coombes, K.R., Tsavachidis, S., Morris, J.S., Baggerly, K.A., Hung, M.C., and Kuerer, H.M. (2004) *Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform*. Technical Report UTMDABTR-001-04, The University of Texas M.D. Anderson Cancer Center.

Diamandis, E.P. (2004) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol.Cell Proteomics. 3(4):367-78.

Eilers, P.H.C. and Marx, B.D. (1996) Flexible smoothing with *B*-splines and penalties, *Statist. Sci.* 11(2): 89–121.

Ezzat, S., Abdel-Hamid, M., Abdel-Latif Eissa S., Mokhtar, N., Labib, N.A., El-Ghorory, L., Mikhail, N.N., Abdel-Hamid, A., Hifnawy, T., Strickland, G.T., and Loffredo, C.A. (2005) Associations of pesticides, HCV, HBV, and hepatocellular carcinoma in Egypt. *Int J Hygiene Env Health* (in press)

Fung, E.T. and Enderwick, C. (2002) ProteinChip clinical proteomics: computational challenges and solutions. *Biotechniques*, 32(suppl), 34-41.

Koomen, J.M., Shih, L.N., Coombes, K.R., Li, D., Xiao, L.C., Fidler, I.J., Abbruzzese, J.L., and Kobayashi, R. (2005) Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins. *Clin Cancer Res.* 11(3):1110-8.

Koopmann, J., Zhang, Z., White, N., Rosenzweig, J., Fedarko, N., Jagannath, S., Canto, M.I., Yeo, C.J., Chan, D.W., and Goggins, M. (2004). Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. *Clin Cancer Res.*, 10 (3), 860-868.

Li, J., Zhang, Z., Rosenzweig, J., Wang, Y.Y., and Chan, D.W. (2002) Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem.* 48 (8), 1296-1304.

Malyarenko, D.I., Cooke, W.E, Adam, B., Malik, G., Chen, H., Tracy, E.R., Trosset, M.W., Sasinowski, M., Semmes, O.J., and Manos, D.M. (2005) Enhancement of Sensitivity and Resolution of Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometric Records for Serum Peptides Using Time-Series Analysis Techniques. *Clinical Chemistry* 51:65-74.

Paradis, V., Degos, F., Dargere, D., Pham, N., Belghiti, J., Degott, C., Janeau, J. L., Bezeaud, A., Delforge, D., Cubizolles, M., Laurendeau, I., and Bedossa, P. (2005) Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases. *Hepatology* 41(1):40-7.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., and Liotta, L.A. (2002a) Use of proteomic patterns in serum to identify ovarian cancer, *Lancet 359*, 572-577.

Petricoin, E.F., Ornstein, D.K., Paweletz, C.P., Ardekani, A., Hackett, P.S., Hitt, B.A., Velassco, A., Trucco, C., Wiegand, L., Wood, K., Simone, C.B., Levine, P.J., Linehan, W.M., Emmert-Buck, M.R., Steinberg, S.M., Kohn, E.C., Liotta, L.A. (2002b) Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst.*, 94 (20), 1576-1578.

Petricoin, E.F., Fishman, D.A., Conrads, T.P., Veenstra, T.D., and Liotta, L.A. (2004) Proteomic pattern diagnostics: producers and consumers in the era of correlative science. *BMC Bioinformatics*. www.biomedcentral.com/1471-2105/4/24/comments

Pusztai, L., Gregory, B.W., Baggerly, K.A., Peng, B., Koomen, J., Kuerer, H.M., Esteva, F.J., Symmans, W.F., Wagner, P., Hortobagyi, G.N., Laronga, C., Semmes, O.J., Wright, G.L. Jr, Drake, R.R., and Vlahou, A. (2004) Pharmacoproteomic analysis of prechemotherapy and postchemotherapy plasma samples from patients receiving neoadjuvant or adjuvant chemotherapy for breast carcinoma. *Cancer* 100(9):1814-1822.

Ransohoff, D.F. (2005) Bias as a threat to the validity of cancer molecular-marker research. Nature Reviews Cancer 5: 142-149.

Ressom H, Varghese RS, Saha D, Orvisky E, Goldman L, Petricoin EF, Conrads TP, Veenstra TD, Abdel-Hamid M, Loffredo CA, and Goldman R. (2005) Particle swarm optimization for analysis of mass spectral serum profiles. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2005*, Beyer, H.G. *et al.*(Editors), ACM Press, New York, vol. 1, pp. 431-438.

Sauve, A.C. and Speed, T.P. (2004) Normalization, baseline correction and alignment of high-throughput mass spectrometry data. Proceedings Gensips 2004.

Semmes, O.J., Feng, Z., Adam, B.L., Banez, L.L., Bigbee, W.L., Campos, D., Cazares, L.H., Chan, D.W., Grizzle, W.E., Izbicka, E., Kagan, J., Malik, G., McLerran, D., Moul, J.W., Partin, A., Prasanna, P., Rosenzweig, J., Sokoll, L.J., Srivastava, S., Srivastava, S., Thompson, I., Welsh, M.J., White, N., Winget, M., Yasui, Y., Zhang, Z., and Zhu, L. (2005) Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem.* 51(1):102-112.

Sorace, J.M. and Zhan, M. (2003) A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* 4 (1), 24.

Villanueva, J., Philip, J., Entenberg, D., Chaparro, C.A., Tanwar, M.K., Holland, E.C., and Tempst, P. (2004) Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal Chem*. 76(6):1560-70.

Yasui, Y., Pepe, M., Thompson, M.L., Adam, B.-L., Wright, G.L. Jr., Qu, Y., Potter, J.D., Winget, M., Thornquist, M., and Feng, Z. (2003) A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics*, 4, 449-463.

Zhang, Z., Bast, R.C. Jr, Yu, Y., Li, J., Sokoll, L.J., Rai, A.J., Rosenzweig, J.M., Cameron, B., Wang, Y.Y., Meng, X.Y., Berchuck, A., Van Haaften-Day, C., Hacker, N.F., de Bruijn, H.W., van der Zee, A.G., Jacobs, I.J., Fung, E.T., and Chan, D.W. (2004) Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res*. 64 (16):5882-5890.

Zhu, W., Wang, X., Ma, Y., Rao, M., Glimm, J., and Kovach, J.S. (2003) Detection of cancer-specific markers amid massive mass spectral data. *Proc Natl Acad Sci* USA, 100 (25), 14666-1467.

RESEARCH ARTICLE

Enrichment of low molecular weight fraction of serum for MS analysis of peptides associated with hepatocellular carcinoma

Eduard Orvisky[1], Steven K. Drake[2], Brian M. Martin[3], Mohamed Abdel-Hamid[4], Habtom W. Ressom[1], Rency S. Varghese[1], Yanming An[1], Daniel Saha[1], Glen L. Hortin[2], Christopher A. Loffredo[1] and Radoslav Goldman[1]

[1] Georgetown University, Department of Oncology, Lombardi Comprehensive Cancer Center, Washington, DC, USA
[2] Clinical Chemistry Service, Department of Laboratory Medicine, NIH, Bethesda, MD, USA
[3] Unit of Molecular Structures, LNT, NIMH, NIH, Bethesda, MD, USA
[4] Viral Hepatitis Research Laboratory, NHTMRI, Cairo, Egypt

A challenging aspect of biomarker discovery in serum is the interference of abundant proteins with identification of disease-related proteins and peptides. This study describes enrichment of serum by denaturing ultrafiltration, which enables an efficient profiling and identification of peptides up to 5 kDa. We consistently detect several hundred peptide-peaks in MALDI-TOF and SELDI-TOF spectra of enriched serum. The sample preparation is fast and reproducible with an average CV for all 276 peaks in the MALDI-TOF spectrum of 11%. Compared to unenriched serum, the number of peaks in enriched spectra is 4 times higher at an S/N ratio of 5 and 20 times higher at an S/N ratio of 10. To demonstrate utility of the methods, we compared 20 enriched sera of patients with hepatocellular carcinoma (HCC) and 20 age-matched controls using MALDI-TOF. The comparison of 332 peaks at $p < 0.001$ identified 45 differentially abundant peaks that classified HCC with 90% accuracy in this small pilot study. Direct TOF/TOF sequencing of the most abundant peptide matches with high probability des-Ala-fibrinopeptide A. This study shows that enrichment of the low molecular weight fraction of serum facilitates an efficient discovery of peptides that could serve as biomarkers for detection of HCC as well as other diseases.

Received: June 14, 2005
Revised: October 18, 2005
Accepted: November 21, 2005

Keywords:
Biomarker / Fibrinopeptide A / Hepatocellular carcinoma / MALDI-TOF MS / Serum

Correspondence: Radoslav Goldman, Georgetown University, LCCC Room S183, 3970 Reservoir Rd NW, Washington, DC 20057-1469, USA
E-mail: rg26@georgetown.edu
Fax: +1-202-6871988

Abbreviations: HCC, hepatocellular carcinoma; **HMW**, high molecular weight; **LMW**, low molecular weight; **SVM**, support vector machine

1 Introduction

Discovery of biomarkers for clinical use typically requires comparison of a large number of samples, which limits the applicability of many elegant proteomic methods [1, 2]. SELDI-TOF and SELDI-QqTOF analyses of serum were optimized for high-throughput comparison of biological samples as small as a few microliters [3]. Advanced statistical

and computational methods were designed to compare the crude mixtures of proteins present in unfractionated serum. The results were surprisingly encouraging given the complexity of the problem and the performance of currently used markers [4–8]. Alpha-fetoprotein, the only serum marker for diagnosis of hepatocellular carcinoma (HCC), has reported sensitivity of 39–64% and specificity of 76–91% [9]. SELDI-based studies of HCC reported sensitivities of 61–90% and specificities of 76–95% [10–13]. Recent analyses raised the question of possible biases in profiling studies, which underscores the need for verification of biomarker identities [14–18]. The identification requires challenging complementary methods in SELDI-TOF experiments [12]. Even SELDI-QqTOF experiments did not sequence the identified biomarker candidates [3]. Here we report an improved protocol that allows an efficient comparison of samples by TOF MS, and identification of many peaks of interest in the low molecular weight (LMW) region by direct TOF/TOF sequencing. The utility of the method was tested in a pilot study of HCC, one of many diseases that could benefit from identification of molecular markers in serum.

2 Materials and methods

2.1 Materials

C8 magnetic bead desalting kits, CHCA, and MALDI 600-μm AnchorChip were purchased from Bruker Daltonics (Billerica, MA, USA). SELDI protein arrays were obtained from Ciphergen (Fremont, CA, USA). Microcon ultrafiltration membranes with 10–50-kDa cut-off were purchased from Millipore (Bedford, MA, USA). Red top vacutainer blood collection tubes (BD 366430) were obtained from Becton Dickinson (Franklin Lakes, NJ, USA). Tricine 10–20% gradient gels for SDS-PAGE and SYPRO Ruby stain were obtained from Invitrogen (Carlsbad, CA, USA). BCA protein assay was purchased from Pierce Biotechnology (Rockford, IL, USA). Other chemicals and solvents were purchased from Sigma-Aldrich (St. Louis, MO, USA); solvents were of HPLC grade.

2.2 Blood samples

A single batch of serum aliquots (standard serum sample) was frozen at −80°C and was used throughout the study to perform method optimization and quality control. Blood samples of cancer cases and controls were obtained between October 2002–April 2003 in collaboration with the National Cancer Institute of Cairo University, Egypt. Controls were recruited among patients from the orthopedic fracture clinic at the Kasr El-Aini Hospital (Cairo, Egypt) and were frequency-matched by gender, rural versus urban birthplace, and age to cancer cases as described previously [19]. Blood samples for all participants were collected around 10 am, were processed within several hours of collection and then immediately frozen at −80°C in 1-mL aliquots. At first thaw, sub-aliquots of 50 μL were generated and stored at −80°C until mass spectrometric analysis. All measurements were performed on aliquots of twice-thawed sera.

2.3 Sample preparation

Serum (15 μL) was desalted on C8 magnetic beads according to the manufacturer's protocol (Bruker Daltonics) and eluted with 50% ACN. Microcon membranes were washed four times with 0.15 mL dH_2O. Samples were diluted with dH_2O to a final concentration of 25% ACN (60 μL total volume) and ultrafiltered at 12 000 × *g* for 5 min using a 50-kDa Microcon membrane. Ultrafiltrate was dried in a centrifugal vacuum evaporator, reconstituted in 2 μL 5% ACN with 0.1% TFA and mixed 1:1 with CHCA matrix (3.3 mg/mL in 50% ACN). The sample (1 μL) was deposited on SELDI gold array or MALDI anchor chip and allowed to crystallize at room temperature.

2.4 SELDI-TOF MS

PBS II Protein Chip Array reader (Ciphergen) was externally calibrated using the $[M+H]^+$ ions of five peptides in the range 1084–7033 *m/z*. All mass spectra were recorded in positive-ion mode at 20 Hz with an 80 ns delay. Saturated solution of CHCA matrix (1 μL) in 50% aqueous ACN containing 0.1% TFA was added to each sample for SELDI-TOF MS analysis.

2.5 MALDI-TOF MS

Samples were analyzed using an Ultraflex MALDI TOF/TOF mass spectrometer (Bruker Daltonics) equipped with pulsed ion extraction ion source. Ionization was achieved by irradiation with a nitrogen laser (λ = 337 nm) operating at 20 Hz. Ions were accelerated at +19 kV with 80 ns of pulsed ion extraction delay. Each spectrum was detected in linear positive mode and was externally calibrated using a mixture of peptide/protein standards between 1 and 4 kDa. Mass spectra were analyzed using the Flex Analysis and ClinProTools softwares (Bruker Daltonics). Raw data were exported as text files for further analysis.

2.6 Gel electrophoresis

Proteins were analyzed by SDS-PAGE using a 10–20% gradient tricine gel and visualized by staining with SYPRO Ruby (Invitrogen) according to manufacturer's protocol.

2.7 Analysis of TOF MS spectra

To reduce the noise and dimensionality of the raw spectra, we used a binning procedure (100-ppm step) that divides the *m/z* axis into 23 846 bins in the 0.9–10 kDa region. The maximum intensity within each interval was used as the

protein expression variable for each spectrum. The baseline of each spectrum was estimated by obtaining the minimum intensity within a shifting widow size of 50 bins. Spline approximation was used to regress the varying baseline and the regressed baseline was subtracted from the spectrum. Each spectrum was normalized by dividing by its total ion current and the spectra were scaled to have an overall maximum intensity of 100. We used two methods for peak detection. One method was used to define S/N ratios and number of peaks [5]. The second method [5] was modified for our study to calculate peak intensities for biomarker discovery. In this latter method, a bin was identified as a peak if the sign of the intensity's slope changed from positive to negative. Peaks with intensity below a pre-defined threshold line were considered as noise and were discarded. To account for drift in m/z location in different spectra, two peaks were coalesced into a window if they differed in location by at most 7 bins or 0.03% relative mass. The maximum intensity in each window was defined as the peak-intensity variable. To distinguish cancer patients and healthy individuals, the processed spectra were split into training and testing (independent) datasets. The training dataset was used to select m/z windows, to compare their intensities using a random variance t-test, and to build a support vector machine (SVM) classifier. Prediction accuracy of the classifier was evaluated using the independent dataset. These functions were performed in BRB-ArrayTools 3.1 software (NCI, Bethesda, MD, USA) [20, 21]. Other analyses were carried out in MATLAB (MathWorks, Natick, MA, USA), Clin ProTools (Bruker Daltonics) and Splus (MathSoft Inc., Cambridge, MA, USA) analytical software packages.

3 Results

Denaturing ultrafiltration enriches the LMW fraction of serum and plasma (Fig. 1). Removal of proteins greater than 50 kDa including albumin appears efficient. SYPRO Ruby staining detected at most traces of albumin in the ultrafiltered serum. Some proteins smaller than 50 kDa are also removed as shown by the SDS-PAGE in Fig. 1. This is consistent with the manufacturer's definition of the 50-kDa cutoff based on retention of analytes. It is expected that partial removal of analytes occurs below the specified cutoff. Our studies focus on peptides <5 kDa because of the biological importance of this fraction of serum and optimal performance of the MALDI-TOF/TOF instrument in this LMW region. Minimal losses are expected in the <5-kDa region provided that protein-protein interactions are disrupted. In this study, ultrafiltration was carried out in the presence of 25% ACN, which allowed removal of high molecular weight (HMW) proteins including albumin with concurrent enrichment of the LMW fraction sufficient for detection of several hundred peptides. Quantification of proteins in the desalted ultrafiltrate by BCA assay (Pierce) showed a 15% increase in the protein content with the addition of

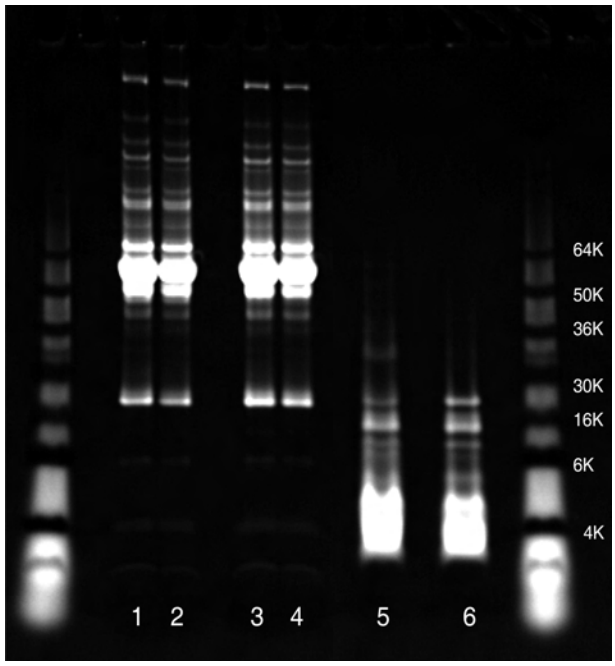


Figure 1. SDS-PAGE analysis of human plasma and serum. Lanes 1 and 2, unfiltered plasma, lanes 3 and 4, unfiltered serum, lane 5, enriched LMW plasma and lane 6, enriched LMW serum; 2 μg total protein was applied per lane and visualized by SYPRO Ruby staining.

ACN (0.30 ± 0.029 μg with 25% ACN; 0.26 ± 0.027 μg without ACN during ultrafiltration). Further experiments are needed to compare various denaturing ultrafiltration conditions [22–24].

The enrichment procedure begins with desalting of serum on C8 magnetic beads, followed by ultrafiltration on a 50-kDa cut-off membrane as described in Materials and methods. The desalted ultrafiltrate can be analyzed on any MS platform of choice. Here we show SELDI-TOF (Fig. 2) and MALDI-TOF (Fig. 3) mass analysis of the LMW fraction spotted with CHCA matrix. The figures align three spectra of independently processed aliquots of a serum standard (out of 9 SELDI spectra and 15 MALDI spectra used to estimate CV). We optimized the method for MALDI-TOF analysis of the 1–5-kDa region using an Ultraflex mass spectrometer (Bruker Daltonics). This instrument allows direct identification of peptides in this mass range by TOF/TOF sequencing.

The method is reproducible as shown by the calculated CV. The mean CV across all peaks (n = 194) in the 9 SELDI-TOF replicate spectra is 17%. The mean CV across all peaks (n = 276) in the 15 MALDI-TOF replicate spectra is 11%. Reproducibility was determined for smoothed, baseline-corrected, and normalized spectra (Fig. 4). Signal processing decreases the CVs approximately two times, which is mostly accounted for by the normalization step. Table 1 compares CVs of small (1–3% maximum peak intensity), medium (5–15% maximum peak intensity), and large (>25% maximum

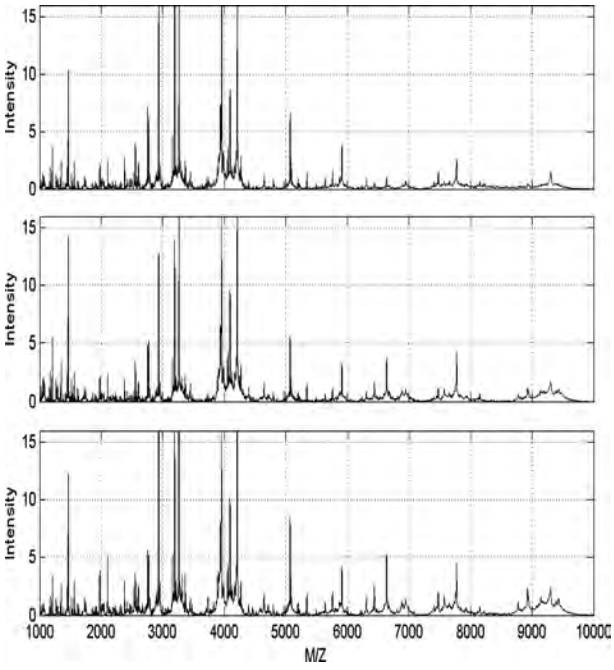


Figure 2. SELDI-TOF spectra, three independently prepared samples of enriched LMW fraction of serum.

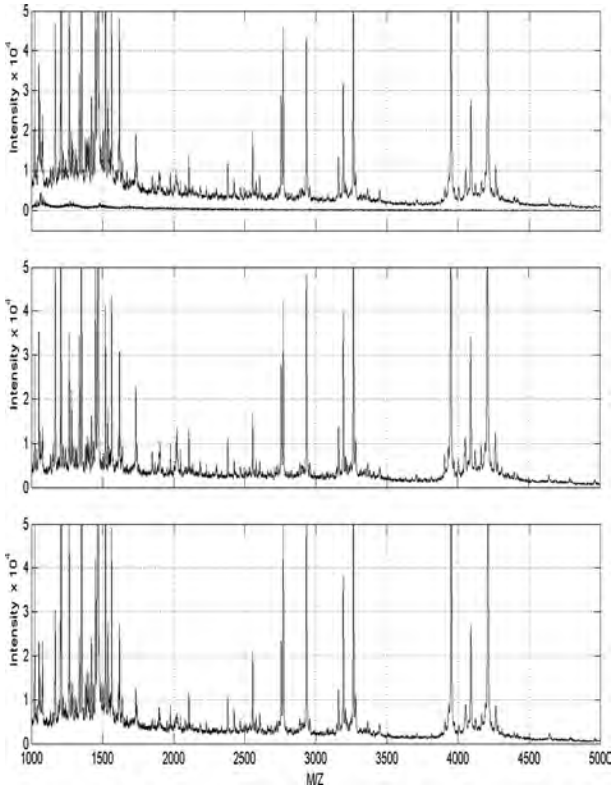


Figure 3. MALDI-TOF spectra, three independently prepared samples of enriched LMW fraction of serum. A spectrum of CHCA matrix without sample is presented for comparison in the top panel.

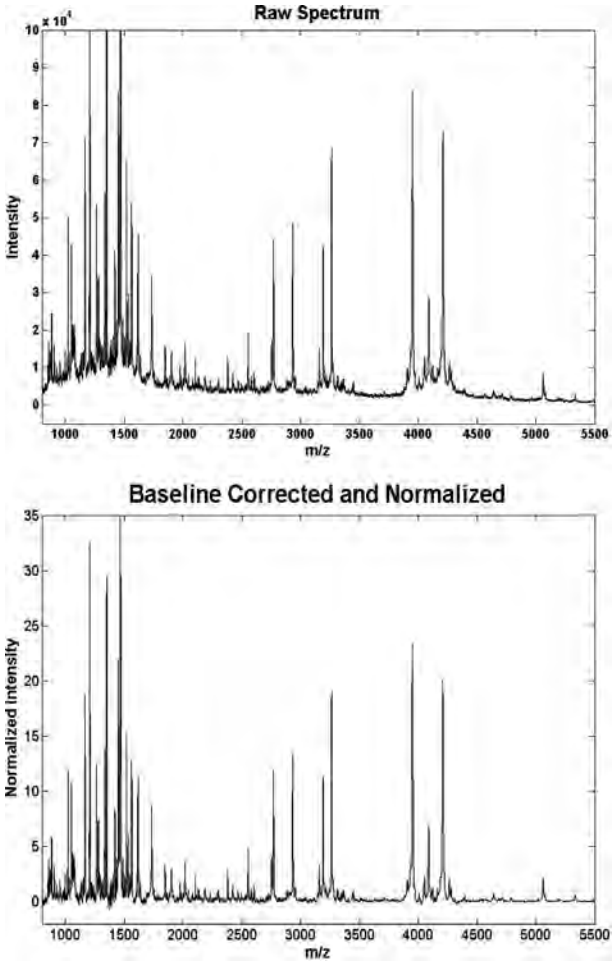


Figure 4. MALDI-TOF spectrum of a standard serum sample processed by smoothing, baseline correction, and normalization. Top panel: raw spectrum; bottom panel: processed spectrum.

Table 1. CV for SELDI-TOF ($n = 9$) and MALDI-TOF ($n = 15$) spectra of independently processed aliquots of standard serum; five peaks were randomLy selected for each group

		CV1	CV2	CV3	CV4	CV5	Mean CV
SELDI	High	9.4	9	16.1	20.6	12.6	13.6
	Medium	17.4	23	29	8.8	11.3	17.9
	Low	18.1	13.2	14.6	14.1	19.3	15.9
MALDI	High	9.3	7.8	7.2	8.9	8.3	8.3
	Medium	12.4	9.3	10.1	13.9	6.8	10.5
	Low	8.8	10.4	9.2	16.8	5.3	10.1

peak intensity) peaks. Five peaks were randomLy selected in each category for this comparison. The mean CV ranges from 8% to 11% for MALDI-TOF and from 14% to 18% for SELDI-TOF spectra in the three intensity groups. The CVs do not vary substantially with signal intensity.

MALDI-TOF spectra of enriched serum prepared as described above have, in our hands, substantially better quality than spectra of C8 desalted serum without ultrafiltration (Fig. 5). The increase in signal quality is demonstrated by a 4-fold increase in the number of peaks at an S/N of 5 and a 25-fold increase at an S/N of 10. It is expected that removal of highly abundant proteins by ultrafiltration under denaturing conditions will improve detection of LMW peptides. We have optimized the procedure such that the eluate of the C8 desalting step is diluted with dH_2O to a final concentration of 25% ACN for the ultrafiltration step. Under these conditions the recovery of peptides from ultrafiltration on the 50-kDa membranes is adequate and reproducible.

The applicability of the method was tested in a pilot study of HCC. We compared spectra of patients with HCC ($n = 20$) and matched controls ($n = 20$) to examine whether we can identify peptides associated with the disease. Overlay of the spectra in ClinProTools software (Bruker Daltonics) is shown in Fig. 6. The picture enlarges the 1200–1900-Da region to better visualize the differences. There are several peaks that strikingly differ between the two groups. This is highlighted in the overlay of average spectra ($n = 20$ for each group) presented in Fig. 7. To carry out a preliminary statistical comparison of this pilot dataset, we first identified peaks in a training set of 10 HCC cases and 10 controls. Our methods defined a total of 332 peaks in the MALDI-TOF spectra yielding a 332×20 matrix of peak intensities. This was used to compare the HCC cases and controls by random variance *t*-test (BRB Array Tools 3.1). The *t*-test identified 45 differentially abundant peaks ($p < 0.001$) that were used to build an SVM classifier. Compared to controls, 34 peaks are increased in HCC and 11 peaks are decreased. The classifier correctly predicted the presence of HCC in 18 of 20 spectra (90% sensitivity and specificity) in an independent test set (10 cases and 10 controls). The independent test spectra were not used for either peak finding or definition of the SVM classifier. Since the test set is small, the data should be viewed as preliminary. The identified peaks are not confirmed biomarkers of HCC and a larger set of samples will have to be used together with multivariate analyses to validate these encouraging pilot results.

The sequence of the most abundant discriminating peptide (*m/z* 1465.6 Da) was defined by TOF/TOF sequencing as DSGEGDFLAEGGGVR, which matches with high probability (MASCOT score 127) the sequence of des-Ala-fibrinopeptide A (Fig. 8). The possibility to directly sequence the peptides of interest is a powerful feature of this method.

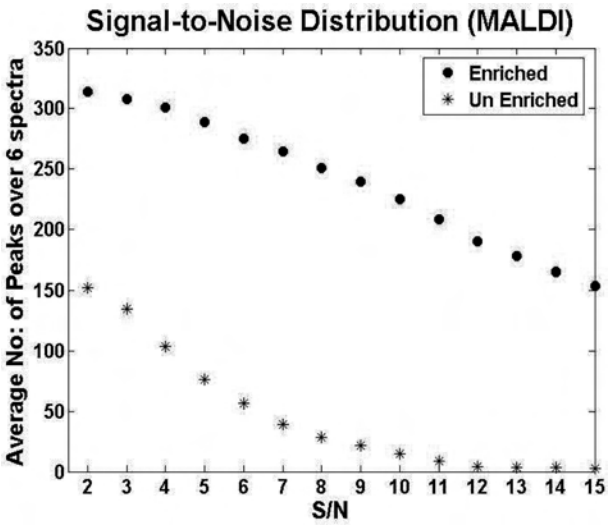


Figure 5. Enrichment of samples by denaturing ultrafiltration expressed as number of peaks at an S/N ratio of 2–15.

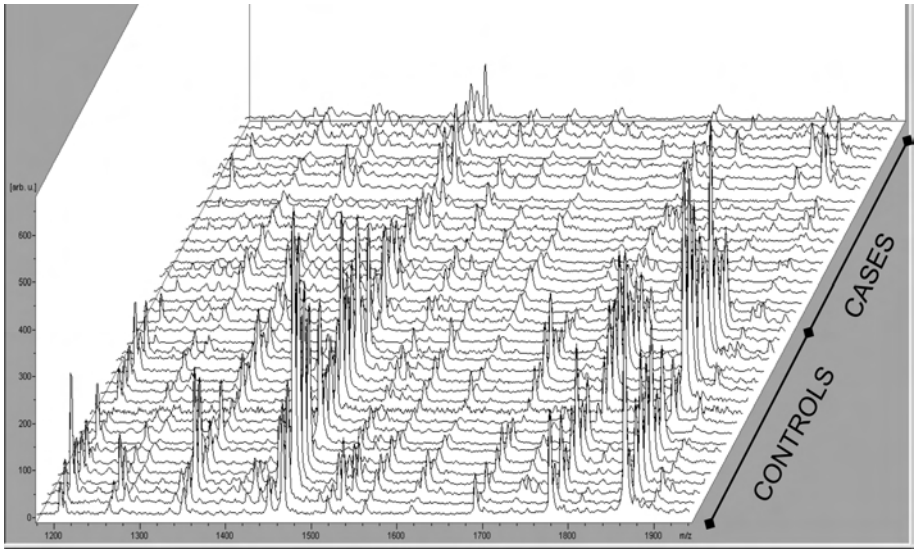


Figure 6. Alignment of 20 HCC cases and 20 controls, mass region of 1200–1900 Da, ClinProTools (Bruker Daltonics).

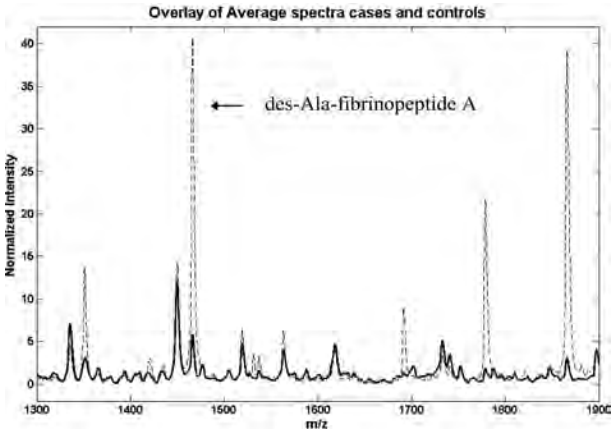


Figure 7. Average spectra of HCC cases and matched controls. Full line: Cases (n = 20); broken line: controls (n = 20).

4 Discussion

Analysis of unfractionated serum samples by SELDI-TOF allows comparison of a large number of samples, but the resolution of the method is limited. It has been proposed that fractionation may be needed to detect less abundant proteins and peptides [17, 25]. More than 95% of serum proteins are represented by 22 species; albumin alone represents about 50% of the serum proteome. Disease-associated proteins are typically found in the remaining fraction [22, 26]. This suggests that removal of HMW carriers without loss of the LMW fraction should improve MS-based biomarker discovery.

We adapted denaturing ultrafiltration, used previously for improved LC-MS/MS and FT-ICR-MS [22, 23, 27], for MALDI TOF/TOF analysis of serum. Our method combines desalting/concentration on C8-derivatized magnetic beads [24, 28] with denaturing ultrafiltration [22]. Desalting on magnetic beads has a higher capacity compared to SELDI surfaces and minimizes handling volumes (compared to small columns, *etc.*). The paramagnetic properties of the particles allow easy handling and automation of the procedure [24]. We selected the 50-kDa cut-off in the presence of ACN, because it efficiently eliminates albumin and other large abundant proteins and allows isolation of an enriched peptide fraction (Figs. 1–5). This method is more effective for isolation of the LMW fraction than immunodepletion, and facilitates comparison of large numbers of samples. It was shown that the abundant HMW proteins bind other peptides/proteins [29–31]. Immunodepletion of the HMW carriers, which can lead to loss of disease-associated peptides, has also been proposed as an enrichment strategy for recovery of the biomarker candidates [32]. Disruption of the interactions by various denaturing conditions including organic solvents is used to improve recovery of peptides [22, 33].

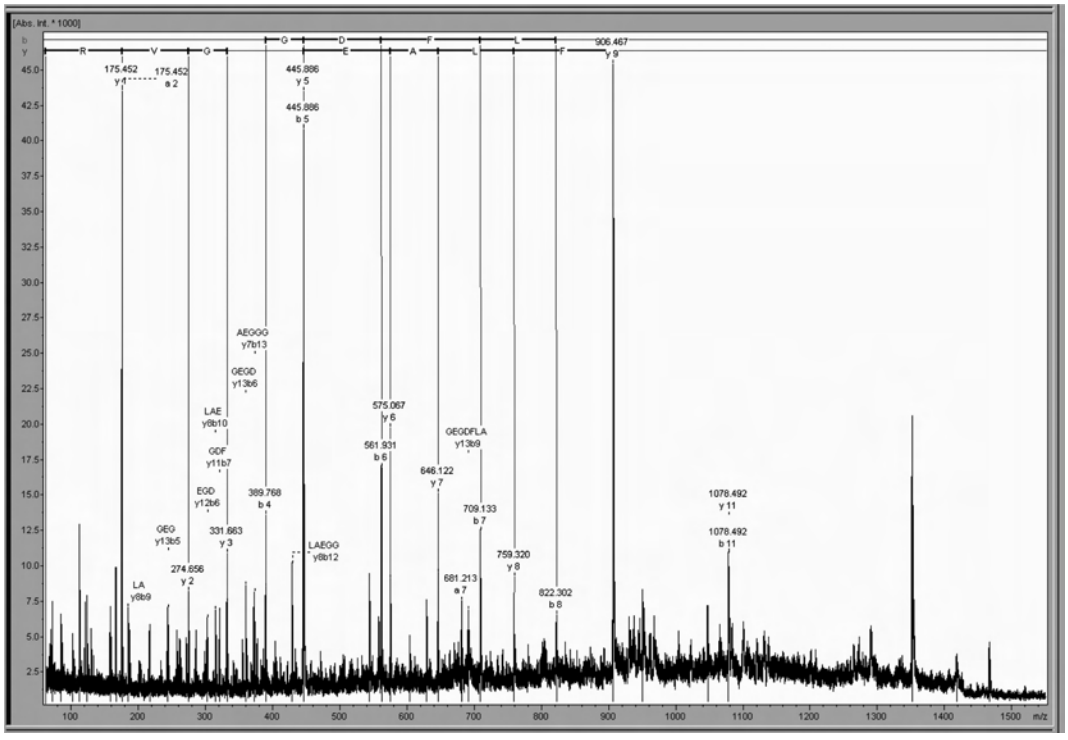


Figure 8. MALDI-TOF/TOF spectrum of peptide with mass 1465.6 Da. Sequence DSGEGDFLAEGGGVR matches with high probability des-Ala-fibrinopeptide A (MASCOT score 127).

The process was streamlined for fast processing of small samples (15 μL serum). In our hands, the enrichment is substantial compared to C8-desalted serum without ultrafiltration (Fig. 5). This method is reproducible with about 10% mean CV across identified peaks (Fig. 5). Reproducibility is critical for biomarker discovery. The enrichment is higher than previously reported likely because of optimal denaturing conditions using ACN and use of different magnetic bead particles [24]. Analysis of unfractionated serum requires careful selection of a batch of magnetic beads with adequate performance [34]. We did observe efficient elimination of albumin and other large proteins [23]. Traces of albumin and presence of other HMW contaminants in our samples did not limit our ability to obtain well-resolved reproducible spectra (Fig. 2). However, we did see elimination of proteins smaller than 50 kDa (Fig. 1). This may be due to short filtration times, which limits contact of ACN with the membrane. Detailed analysis of filters with lower cut-off under various denaturing conditions is needed to further optimize the enrichment for specific applications.

Our method was applied to a pilot analysis of HCC, one of many diseases that could benefit from improved classification based on molecular markers in serum. HCC is a common cancer worldwide with as many as 500,000 new cases each year [35]. Egypt is a country with high rates of chronic HCV infection and associated HCC [36, 37]. The 40 samples analyzed in this study are a subset of a large study of HCC we conducted in Egypt [19]. We proposed that a set of peptides associated with HCC could be present in serum and serve as a biomarker. Cancer biomarkers in serum are not necessarily new antigens; they are often modified peptides and fragments of proteins [22, 26, 38]. Proteases and peptidases are reportedly deregulated in HCC [39–41], and it is reasonable to expect that cancer-specific fragments of proteins will be found in the LMW region [12].

Here we demonstrate that our method is efficient for discovery of peptides associated with HCC. We should emphasize that the described peptides are not validated biomarkers. Larger studies and additional analyses are needed to confirm that these biomarker candidates have utility. Nonetheless, the results are encouraging. We identified 45 peptides that are associated with HCC and classify the disease with 90% prediction accuracy in this small pilot. Importantly, the method allows direct identification of the peptides by TOF/TOF sequencing as demonstrated by the identification of the peak at m/z 1465.6 as a fragment of fibrinopeptide A.

Recent studies noted that human serum contains fragments of relatively common proteins [42]. These fragments are expected to be present in the ultrafiltered serum under our experimental conditions. A fragment of fibrinopeptide A was in fact the most abundant peptide in our study (Fig. 7) and in a comparable study of ultrafiltered serum using FT-ICR [27]. It is not clear whether these fragments and their PTMs can be used for disease classification. The specificity of these fragments and modified peptides was not explored. The composition of the peptide mixture in our samples remains to be defined. The distinct pattern of peptides in HCC is most likely related to differential proteolytic activity in cancer patients. It will be important to define whether the peptides represent cancer-related antigens, fragments derived from the activity of tumor-related proteases, or host response. The use of combinations of peptides to define disease status was not studied extensively. Currently, it is unclear whether a combination of peptides related to tumor-related enzymatic activities or host response can provide an efficient biomarker.

In summary, we describe a sensitive high-throughput platform for discovery of biomarker candidates among peptides in the LMW fraction of serum. This method combines C8 desalting and denaturing ultrafiltration for simultaneous measurement of several hundred peptides by MALDI-TOF. Biomarkers can improve disease classification, early detection and intervention, assessment of disease progression, and possibly long-term outcomes. It will be important to use all available information (quantity, sequence, modifications, etc) to find optimal peptide biomarkers and their combinations. The presented method is expected to facilitate the discovery of biomarkers among peptides in the LMW fraction of serum.

This work was supported in part by U.S. Army Medical Research and Material Command, Prostate Cancer Research Program grant DAMD17-02-1-0057, NIEHS grant 1R21ES011958-01A1, and American Cancer Society grant CRTG-02-245-01-CCE awarded to R.G. We want to acknowledge Drs. Soheir Abdel-Latif Eissa and Alaa Ismail for invaluable help with design and execution of the studies in Egypt. We thank Drs. F. Seillier-Moiseiwitsch and A. Wang for their helpful suggestions and discussions.

5 References

[1] Veenstra, T. D., Conrads, T. P., Hood, B. L., Avellino, A. M. *et al.*, *Mol. Cell. Proteomics* 2005, *4*, 409–418.

[2] Hanash, S., *Nature* 2003, *422*, 226–232.

[3] Conrads, T. P., Fusaro, V. A., Ross, S., Johann, D. *et al.*, *Endocr. Relat Cancer* 2004, *11*, 163–178.

[4] Petricoin, E. F., Zoon, K. C., Kohn, E. C., Barrett, J. C. *et al.*, *Nat. Rev. Drug Discov.* 2002, *1*, 683–695.

[5] Coombes, K. R., Fritsche, H. A., Jr., Clarke, C., Chen, J. N. *et al.*, *Clin. Chem.* 2003, *49*, 1615–1623.

[6] Qu, Y., Adam, B. L., Thornquist, M., Potter, J. D. *et al.*, *Biometrics* 2003, *59*, 143–151.

[7] Ressom, H. W., Varghese, R. S., Abbel-Hamid, M., Eissa, S. A. *et al.*, *Bioinformatics* 2005, *21*, 4039–4045.

[8] Yasui, Y., Pepe, M., Thompson, M. L., Adam, B. L. *et al.*, *Biostatistics* 2003, *4*, 449–463.

[9] Lopez, L. J., Marrero, J. A., *Curr. Opin. Gastroenterol.* 2004, *20*, 248–253.

[10] Schwegler, E. E., Cazares, L., Steel, L. F., Adam, B. L. *et al.*, *Hepatology* 2005, *41*, 634–642.

[11] Zhu, X. D., Zhang, W. H., Li, C. L., Xu, Y. *et al.*, *World J. Gastroenterol.* 2004, *10*, 2327–2329.

[12] Paradis, V., Degos, F., Dargere, D., Pham, N. *et al.*, *Hepatology* 2005, *41*, 40–47.

[13] Poon, T. C., Yip, T. T., Chan, A. T., Yip, C. *et al.*, *Clin. Chem.* 2003, *49*, 752–760.

[14] Liotta, L. A., Lowenthal, M., Mehta, A., Conrads, T. P. *et al.*, *J. Natl. Cancer Inst.* 2005, *97*, 310–314.

[15] Petricoin, E. F., Fishman, D. A., Conrads, T. P., Veenstra, T. D. *et al.*, *Proteomics* 2004, *4*, 2357–2360.

[16] Baggerly, K. A., Morris, J. S., Coombes, K. R., *Bioinformatics* 2004, *20*, 777–785.

[17] Diamandis, E. P., *Mol. Cell. Proteomics* 2004, *3*, 367–378.

[18] Ransohoff, D. F., *J. Natl. Cancer Inst.* 2005, *97*, 315–319.

[19] Ezzat, S., Abdel-Hamid, M., Eissa, S. A. L., Mokhtar, N. *et al.*, *Int. J. Hyg. Environ. Health* 2005, *208*, 329–339.

[20] Simon, R., Radmacher, M. D., Dobbin, K., McShane, L. M., *J. Natl. Cancer Inst.* 2003, *95*, 14–18.

[21] Simon, R., *Br. J. Cancer* 2003, *89*, 1599–1604.

[22] Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J. *et al.*, *Mol. Cell. Proteomics* 2003, *2*, 1096–1103.

[23] Harper, R. G., Workman, S. R., Schuetzner, S., Timperman, A. T. *et al.*, *Electrophoresis* 2004, *25*, 1299–1306.

[24] Villanueva, J., Philip, J., Entenberg, D., Chaparro, C. A. *et al.*, *Anal. Chem.* 2004, *76*, 1560–1570.

[25] Ransohoff, D. F., *J. Natl. Cancer Inst.* 2005, *97*, 315–319.

[26] Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S. *et al.*, *Proteomics* 2003, *3*, 1345–1364.

[27] Bergen, H. R. III, Vasmatzis, G., Cliby, W. A., Johnson, K. L. *et al.*, *Dis. Markers* 2003, *19*, 239–249.

[28] Zhang, X., Leung, S. M., Morris, C. R., Shigenaga, M. K., *J. Biomol. Tech.* 2004, *15*, 167–175.

[29] Zhou, M., Lucas, D. A., Chan, K. C., Issaq, H. J. *et al.*, *Electrophoresis* 2004, *25*, 1289–1298.

[30] Liotta, L. A., Ferrari, M., Petricoin, E., *Nature* 2003, *425*, 905.

[31] Lowenthal, M. S., Mehta, A. I., Frogale, K., Bandle, R. W. *et al.*, *Clin. Chem.* 2005, *51*, 1933–1945.

[32] Mehta, A. I., Ross, S., Lowenthal, M. S., Fusaro, V. *et al.*, *Dis. Markers* 2003, *19*, 1–10.

[33] Fu, Q., Garnham, C. P., Elliott, S. T., Bovenkamp, D. E. *et al.*, *Proteomics* 2005, *5*, 2656–2664.

[34] Villanueva, J., Philip, J., Chaparro, C. A., Li, Y. *et al.*, *J. Proteome Res.* 2005, *4*, 1060–1072.

[35] Montalto, G., Cervello, M., Giannitrapani, L., Dantona, F. *et al.*, *Ann. N. Y. Acad. Sci.* 2003, *963*, 13–20.

[36] Kew, M. C., *Clin. Lab Med.* 1996, *16*, 395–406.

[37] Nada, O., Abdel-Hamid, M., Ismail, A., El Shabrawy, L. *et al.*, *J. Clin. Virol.* 2005, *34*, 140–146.

[38] Schulz-Knappe, P., Zucht, H. D., Heine, G., Jurgens, M. *et al.*, *Comb. Chem. High Throughput Screen.* 2001, *4*, 207–217.

[39] Zhou, X. D., *Hepatobiliary Pancreat. Dis. Int.* 2002, *1*, 35–41.

[40] Sun, Z., Yang, P., *Lancet Oncol.* 2004, *5*, 182–190.

[41] Feitelson, M. A., Pan, J., Lian, Z., *Surg. Clin. North Am.* 2004, *84*, 339–354.

[42] Koomen, J. M., Shih, L. N., Coombes, K. R., Li, D. *et al.*, *Clin. Cancer Res.* 2005, *11*, 1110–1118.

Analysis of MALDI-TOF Serum Profiles for Biomarker Selection and Sample Classification

H. W. Ressom[*1], R. S. Varghese[1], E. Orvisky[1], S. K. Drake[2], G. L. Hortin[2],

M. Abdel-Hamid[3], C. A. Loffredo[1], and R. Goldman[1]

[1]Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC
[2]Clinical Chemistry Service, Department of Laboratory Medicine, NIH, Bethesda, MD
[3]Viral Hepatitis Research Laboratory, NHTMRI, Cairo, Egypt
[*]Corresponding author: hwr@georgetown.edu

***Abstract*- Mass spectrometric profiles of peptides and proteins obtained by current technologies are characterized by complex spectra, high dimensionality, and substantial noise. These characteristics generate challenges in discovery of proteins and protein-profiles that distinguish disease states, e.g. cancer patients from healthy individuals. A challenging aspect of biomarker discovery in serum is the interference of abundant proteins with identification of disease-related proteins and peptides. We present data processing methods and computational intelligence that combines support vector machines (SVM) with particle swarm optimization (PSO) for biomarker selection from MALDI-TOF spectra of enriched serum. SVM classifiers were built for various combinations of m/z windows guided by the PSO algorithm. The method identified mass points that achieved high classification accuracy in distinguishing cancer patients from non-cancer controls. Based on their frequency of occurrence in multiple runs, six m/z windows were selected as candidate biomarkers. These biomarkers yielded 100% sensitivity and 91% specificity in distinguishing liver cancer patients from healthy individuals in an independent dataset.**

I. INTRODUCTION

Mass spectrometric serum profiling was optimized for high-throughput comparison of complex samples that allows discovery of biomarkers of diseases such as cancer [1]. Independent analysis of the results pointed out the importance of avoiding bias and the need for independent validation of results [2-4]. Improved study design and technology in second-generation studies continue to indentify biomarker-candidates for variety of cancers [5-7]. This paper adds data preprocessing and feature selection methods to a growing number of improved tools for matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometric identification of biomarkers in enriched serum.

Mass spectra represent a complex signal consisting of electronic noise, chemical noise due to contaminants and matrix, and protein and metabolic signatures [8]. They also have a varying baseline caused, besides others, by matrix-associated chemical noise or by ion overload. The latter refers to the high excess of ions derived from the matrix that can overload the detector [9]. This elevates the baseline from its ideal zero horizontal line. Previous quality-control experiments have suggested several measurement properties of current mass spectrometry technologies that must be accounted for in the analysis [10]. These properties include high dimensionality of the spectra, high coefficients of variation, and mass shift (measurement error). Thus, it is important to apply preprocessing methods that enable the recognition of spectral quality prior to using the spectra for biomarker discovery and sample classification.

Data preprocessing methods such as smoothing, baseline correction, normalization, peak detection, and peak alignment improve the performance of mass spectrometric data analysis methods for biomarker discovery [9, 11]. The reason for this includes the substantial amount of noise and systematic variations between spectra caused by sample degradation over time, ionization suppression, and other parameters reviewed previously by [4, 12]. The data preprocessing methods are typically available in all software for operation of a mass spectrometer. The use of spectral comparisons for biomarker identification requires, however, optimization of these methods and a completely transparent data manipulation. Several groups proposed recently improved tools for data preprocessing and biomarker discovery as summarized briefly below.

By smoothing the raw spectra, we can reduce the effect of some mass-per-charge (m/z) values that appear as peaks but may not be or are very hard to verify by independent experiments. Many smoothing algorithms are available to denoise raw signals including the well-known Savitzky-Golay filter that removes additive white noise [13] and wavelets [14].

Baseline correction is important for minimization of background noise; drifting baseline introduces serious distortion of ion intensities without adequate correction. Several methods have been proposed for baseline subtraction. For example, Fung and Enderwick [15] employed a varying-width segemented convex hull algorithm to subtract the baseline. Baggerly *et al.* [16] fitted a local median or local mean in a fixed window on the time scale. They also considered subtracting a "semimonotonic" baseline. Coombes *et al.* [14] estimated baseline by fitting a monotone local minimum curve to smoothed spectra.

Normalization reduces variation in signal intensity between spectra. A commonly used normalization method for mass spectrometric data is rescaling each spectrum by its total ion current, i.e., the area under the curve (AUC) [11, 15]. Other common choices for the rescaling coefficient include the

spectrum median or mean. Alternatively, choosing the average AUC over all spectra as the rescaling coefficient can do a global normalization. A global optimization assumes that the sample intensities are all related by a constant factor. That means that the data distribution should not differ substantially from one spectrum to another.

Peak detection deals with the selection of m/z values which display a reasonable intensity compared to those that appear as noise. Coombes *et al.* [14] applied a simple peak finding (SPF) algorithm that provides the locations of potential peaks and their associated left-hand and right-hand bases. They estimated signal-to-noise ratio (S/N) using wavelets for improved peak detection. Also, they introduced a method for coalescing neighboring peaks.

Assuming appropriate mass spectral data preprocessing methods are used, biomarker selection can be addressed using various computational methods. One of the commonly used approaches is to apply statistical analyses that recognize differentially expressed m/z values between cases and controls with multiple subjects. For example, one can apply a two-sample t-test method to compare the protein intensities at each m/z value in cases and controls. Zhu *et al.* [17] proposed a statistical algorithm that can select a subset of *k* biomarkers from the marker list that could best discriminate between the groups in a training dataset via the best *k*-subset discriminant method with high sensitivity and specificity.

Computational intelligence has also been applied for biomarker discovery. For example, Petricoin *et al.* [1] used a combination of genetic algorithm (GA) and self-organizing clustering (GA-SOC) for variable selection. The GA-SOC, which is implemented in ProteomeQuest software, starts with hundreds of random choices of small sets of exact m/z values selected from the SELDI-TOF mass spectra. Each candidate subset contains 5 to 20 of the potential m/z values that define the spectra. The m/z values within the highest rated sets are reshuffled to form new subset candidates. The candidates are rated iteratively until the set that fully discriminates the preliminary set emerges.

Koopmann *et al.* [18] applied successfully support vector machines (SVMs) in a modified form to proteomic profiling. Li *et al.* (2002) introduced unified maximum separability analysis (UMSA) algorithm, which incorporates data distribution information into structural risk minimization learning algorithm. UMSA is applied to identify a direction along which two classes of data are best separated. This direction is represented as a linear combination of the original variables. The weight assigned to each variable in this combination measures the contribution of the variable toward the separation of the two classes of data. They analyzed protein profiles of serum samples from patient with or without breast cancer. They reported that UMSA enabled the identification of three discriminatory biomarkers that achieved 93% sensitivity and 91% specificity in detecting breast cancer patients from the non-cancer controls.

In our previous work [19, 20], we proposed a novel computational method known as PSO-SVM that combines SVMs and particle swarm optimization (PSO) for optimal selection of m/z values from high resolution surface enhanced laser desorption ionization-quadrupole time-of-flight (SELDI-QqTOF) spectra. In [20], we performed binning, normalization, baseline correction, peak identification, and peak alignment on SELDI-QqTOF spectra. The peak alignment method defines windows of m/z values that have variable width. The PSO-SVM algorithm is then applied to select the optimal m/z windows. We ran the algorithm multiple times and selected five m/z windows based on their frequency of occurrence. An SVM classifier that employs these five m/z windows as its inputs yielded 92% sensitivity and 90% specificity in distinguishing hepatocellular carcinoma (HCC) patients from matched controls.

In this paper, the serum samples were enriched by denaturing ultrafiltration and desalting [21] on C8 magnetic beads (MB) [22]. The procedure disrupts protein-protein interactions and allows an efficient recovery of a low molecular weight (LMW) serum fraction starting with 25 µl of serum. The enrichment offers more peaks than unenriched SELDI-QqTOF or unenriched MALDI-TOF spectra [23]. This paper presents our analysis of MALDI-TOF spectra of enriched serum for biomarker discovery and sample classification.

II. Methods and Results

A. Mass Spectral Data

The incidence of HCC in the United States increases. Very high rates of HCC incidence are observed in Egypt where an epidemic of viral infections presents a serious health problem. The management of the disease would benefit from identification of biomarkers related to this disease. Serum samples of HCC cases and controls were obtained from 2000 to 2002 in collaboration with the National Cancer Institute of Cairo University, Egypt. Controls were recruited among patients from the orthopedic fracture clinic at the Kasr El-Aini Hospital, Cairo, Egypt and were frequency-matched to cancer cases by gender, rural versus urban birthplace, and age [24]. Blood samples were collected by trained phlebotomist each day around 10am and processed within a few hours according to a standard protocol. Aliquots of sera for mass spectrometric analysis were frozen at -80^{o}C immediately after collection till analysis; all measurements were performed on samples of second-time thawed serum.

Eluted peptides in the enriched serum samples were mixed with a matrix solution (3 mg/ml α-cyano-4-hydroxycinaminic acid in 50% actonitrile with 0.1% trifluoracetic acid), spotted onto AnchorChip target (Bruker Daltonics, Billerica, MA), and analyzed using an Ultraflex MALDI TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA). Each spectrum was detected in linear positive mode and was externally calibrated using a standard mixture of peptides. Denaturing ultrafiltration enriches LMW fraction of serum and plasma (Fig. 1). Removal of proteins greater than 50 kDa including albumin appears complete based on Coomassie staining; proteins smaller than 50kDa are also removed as shown by the SDS-PAGE in Fig. 1 (left). Fig. 1 (right) depicts the spectrum found after desalting (top spectrum) and after

denaturing ultrafiltration (bottom spectrum). The enrichment improved the quality of the spectrum in the LMW region and provided over 300 peaks. Evidently, the enrichment (bottom spectrum) offers more peaks than an unenriched spectrum (top spectrum).

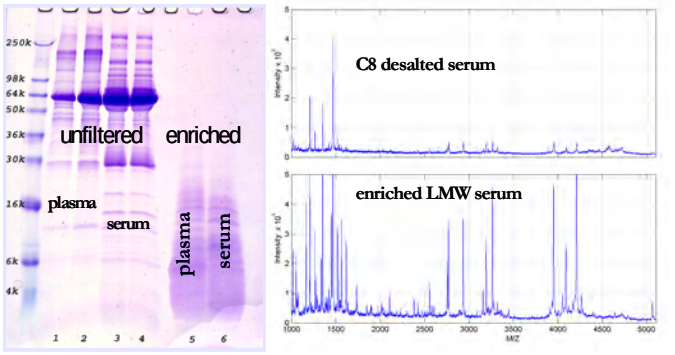


Fig. 1. *Left:* SDS-PAGE analysis of human plasma and serum. *Lane 1 and 2*, unfiltered plasma, *Lanes 3 and 4*, unfiltered serum, *lane 5*, enriched LMW plasma and *lane 6*, enriched LMW serum. 10 µg of total protein was applied per lane and visualized by Coomassie staining. *Right*: MALDI-TOF spectrum after desalting using C8 magnetic beads (top spectrum) and after denaturing ultrafiltration. (bottom spectrum).

B. Reproducibility

Our study used 62 replicate spectra to examine the reproducibility of MALDI-TOF mass spectrometry. Each spectrum consisted of ~136,000 m/z values with the corresponding ion intensities. The dimension of these high-resolution spectra was reduced to 23,846 m/z values using the binning procedure that divides the m/z axis into intervals of desired length over the mass range 0.9 to 10 kDa. A bin size of 100 parts per million (ppm) was found adequate. The mean of the intensities within each interval was used as the protein expression variable in each bin. Each intensity value was transformed by computing the base-two logarithm and found the mean log intensity value and standard deviation.

The CV of the log-transformed intensity values in the 62 reference spectra ranged between 4.1% and 22.9% with mean value of 10.5%. A heat map for 62 replicate spectra (Fig. 2) and spectra for three independently prepared samples of enriched LMW fraction of serum (Fig. 3) illustrate the reproducibility of MALDI-TOF MS.

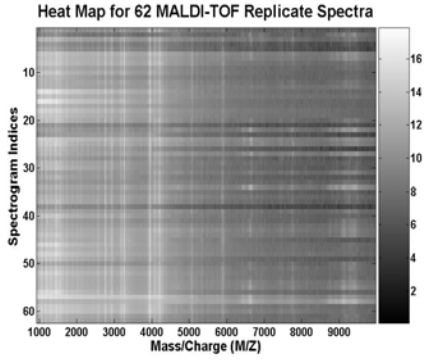


Fig 2. Heat map for 62 MALDI-TOF replicate spectra.

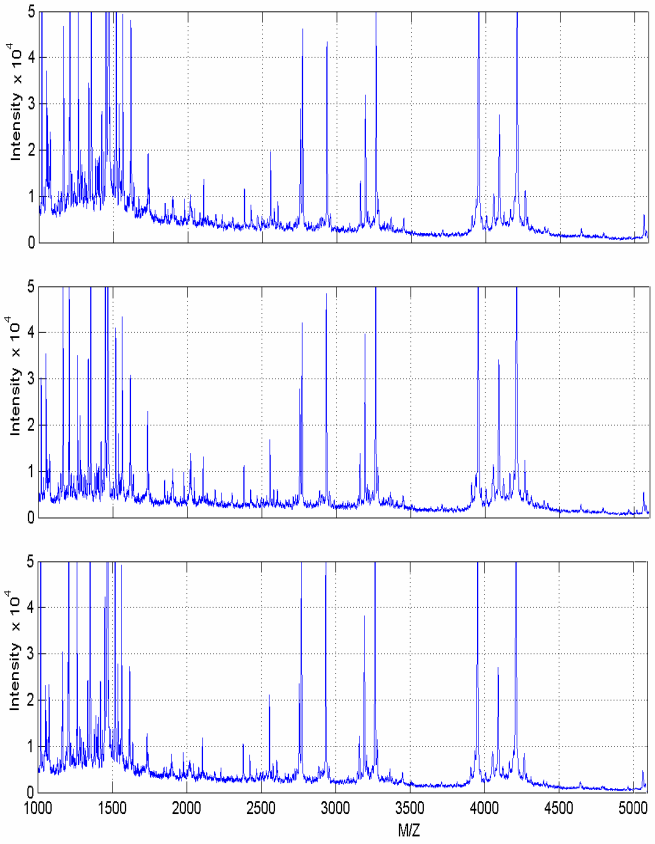


Fig 3. MALDI-TOF spectra, three independently prepared samples of enriched LMW fraction of serum.

C. Data Preprocessing

We applied various methods to preprocess the raw MALDI-TOF mass spectra. We began our analysis with outlier screening where spectra whose data distribution substantially deviated from others were removed. 14 of the 164 MALDI-TOF spectra were excluded, leaving 150 (78 cases and 72 controls) serum mass spectral profiles for further analysis. These outliers were singled out based on their deviation from the median ion current, median record count (number of mass points), and their alignment with pre-selected landmarks.

Each spectrum was first binned with a bin size of 100 ppm, which reduced the dimension of the spectra from about 136,000 m/z values to 23,846 bins over the mass range 0.9 to 10 kDa. Figure 4a and 4b depict a typical raw spectrum of a healthy individual and the corresponding binned spectrum, respectively. On the horizontal axis are m/z values or bins and on the vertical axis are intensity measurements that indicate the relative ion abundance. As shown in the figures, the binning algorithm has removed some high frequency noise, thus smoothing the spectrum. Also, binning improves the alignment of multiple spectra (not shown).

The baseline of each binned spectrum was estimated by obtaining the minimum value within a shifting window size of 50 bins. Spline approximation was used to regress the varying baseline. The regressed baseline was subtracted from the spectrum yielding a baseline corrected spectrum. Spline

regression estimates different linear slopes for different ranges of the m/z values. Eilers and Marx [25] applied the method for baseline correction of 2-D gel electrophoresis images. Furthermore, each spectrum was normalized by dividing it by its total ion current. In addition, the spectra were scaled to have an overall maximum intensity of 100. Fig. 4c shows the binned, normalized, and baseline corrected spectrum.

For peak detection, a bin is identified as a peak if the sign of the intensity's slope changes from positive to negative. Peaks with intensity below a pre-defined threshold-line were considered as noise and were discarded. We selected m/z values with reasonable intensity levels and discarded those that appeared as noise. Following outlier screening, binning, baseline correction, normalization, and peak detection, the 78 HCC case and 72 control spectra were split into 100 training spectra (50 HCC and 50 normal) and 50 testing spectra (28 HCC and 22 normal). The testing spectra were scaled based on the parameters used for scaling the training spectra.

To account for variation in the m/z location (drifts) in different spectra, two peaks were coalesced if they differed in location by at most 7 bins or at most 0.03% relative mass. This method was based on the ideas of Coombes *et al.* [14] who used this method for SELDI-TOF spectra, where they combined peaks if they fall within 7 clock ticks and differ by at most 0.3% relative mass. We applied this method on training dataset only and found 264 windows. Fig. 5 shows m/z windows found between 1730 and 1870 Da. For each spectrum, the maximum intensity within each window was found, yielding a 264 x 100 training data matrix. The same windows were used to quantify the peaks in the testing spectra, which resulted in a 264 x 50 testing data matrix.

D. PSO-SVM

The PSO-SVM algorithm can be used to identify optimal m/z windows from preprocessed mass spectra. While PSO selects subsets of predefined m/z windows as potential solutions, SVM classifiers are built for each potential solution generated by PSO. The prediction capability of the resulting SVM classifier on a validation dataset is used as a performance function for the PSO algorithm. Since SVMs provide good generalization capability in classification tasks and can be designed in a computationally efficient manner, they are an ideal candidate for use as a performance function.

The training dataset is used to select m/z windows and build an SVM classifier. The validity of each classifier trained with the selected features is evaluated using the prediction accuracy of the SVM classifier in distinguishing cancer patients from non-cancer controls. SVM classifiers are built for various combinations of features until the performance of the SVM classifier converges or a pre-specified maximum iteration number is reached.

Estimates of prediction accuracy are calculated by using the *k*-fold cross-validation and bootstrapping methods. In *k*-fold cross-validation, we divide the training dataset into *k* subsets of (approximately) equal size. We train the SVM classifier *k* times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the prediction accuracy.

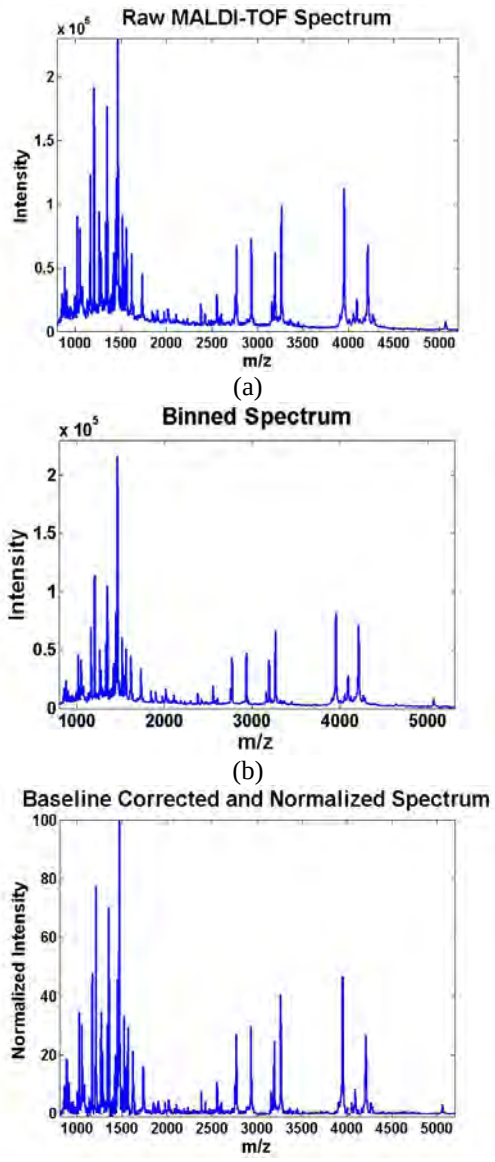


Fig. 4. MALDI-TOF spectrum of a standard serum sample processed by smoothing, baseline correction, and normalization. (a) raw ; (b) binned; and (c) binned, normalized, and baseline corrected spectrum.

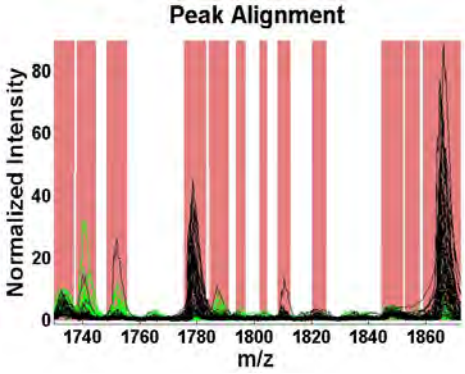


Fig. 5. Control spectra (black), case spectra (light), windows in the m/z range from 1.73 to 1.87 kDa.

In bootstrapping, instead of analyzing pre-specified subsets of the training dataset, we repeatedly select subsamples of the data. Each subsample is a random sample with replacement from the full training dataset.

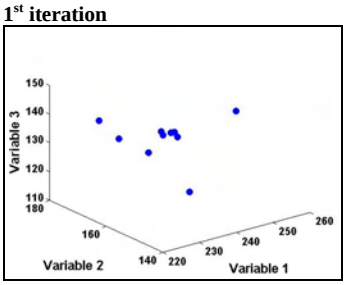
The PSO-SVM algorithm is used to identify the optimal m/z windows from a list of *L* potential m/z windows. The algorithm creates *N* vectors (particles), each consisting of *n* m/z windows that are randomly selected from *L* m/z windows. The algorithm evaluates the performance of each particle in distinguishing cancer cases from controls. This is carried out by building an SVM classifier for each particle and evaluating the performance of the classifier via the *k*-fold cross-validation or bootstrapping methods. The algorithm uses the most-fit particles to contribute to the next generation of *N* candidate particles. Thus, on the average, each successive population of candidate particles fits better than its predecessor. This process continues until the performance of the SVM classifier converges.

The algorithm repeats the above steps multiple times and provides a list of selected m/z windows along with their frequency of occurrence. A frequency plot is used to estimate the optimal number of m/z windows. The frequency plot presents the number of occurrences versus the m/z windows sorted in the order of decreasing frequency. We considered as candidate biomarkers all m/z windows starting from the first until the frequency curve becomes flat (i.e. the change in frequency becomes low). These m/z windows are evaluated via testing dataset (i.e., independent dataset that was used neither for training nor for variable selection) to determine the generalization capability of the SVM classifier.

We present as an example a single run to demonstrate how the PSO-SVM algorithm selects three markers (n=3) out of 264 m/z windows (*L*=264) using 100 MALDI-TOF spectra. The number of particles in this example is 10 (*N*=10). Note that the algorithm searches in a continuous search space but the numbers are rounded to the nearest integer. The elements of a particle represent the variable set suggested by the particle. Each particle is used to build an SVM classifier. In this example, the performance of the SVM classifier is evaluated through the bootstrapping method that randomly splits the spectra (80% for building an SVM classifier and the remaining 20% for validation). This is repeated 500 times with resubstitution and the average prediction accuracy on the validation set is computed.

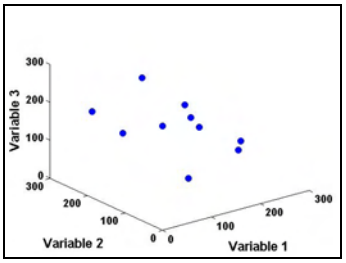
Fig. 6 shows the variable sets selected and their prediction accuracy on the validation set at the 1st, 100th, and 500th iterations, respectively. The left panel depicts the location of the particles in a three-dimensional space. The tables in the right panel show the corresponding coordinates sorted in decreasing order of their prediction accuracy (only the top three and the bottom two variable sets among the 10 variable sets are presented). As shown in the figure, the particles converged to one location (240, 162, 135) after 500 iterations improving the prediction accuracy from 77% to 91%. This location corresponds to m/z windows 4644.9-4651.4, 2528.7-2535.5, and 1863.4-1871.3.

1st iteration



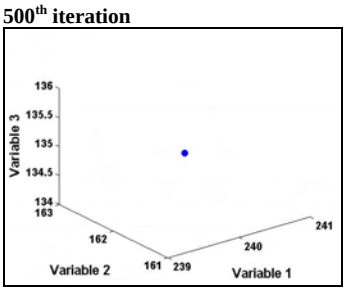
Selected Variable Sets			Accu. %
239	213	90	77
156	257	230	73
75	25	139	72
.	.	.	.
.	.	.	.
99	234	115	60
172	224	112	54

100th iteration



Selected Variable Sets			Accu. %
240	162	135	91
239	162	135	90
237	162	135	85
.	.	.	.
.	.	.	.
227	144	146	63
228	173	138	58

500th iteration



Selected Variable Sets			Accu. %
240	162	135	91
240	162	135	91
240	162	135	91
.	.	.	.
.	.	.	.
240	162	135	91
240	162	135	91

Fig. 6. Variable sets selected by the PSO-SVM algorithm and their prediction accuracy at the 1st, 100th, and 500th iterations. The figures in the left panel show the location of the particles in the three-dimensional space. Each table in the right panel shows the top three and the bottom two variable sets among the 10 variable sets (particles) used by PSO, sorted in decreasing order of prediction accuracy.

E. Biomarker Selection

The purpose of this analysis is to identify optimal m/z windows or candidate biomarkers from the preprocessed mass spectral data. While peak detection deals with the selection of mass points with reasonable intensity and S/N ratio, the aim of biomarker selection is to identify mass points that can be used to distinguish between cancer patients and healthy individuals.

We used the PSO-SVM algorithm to select candidate biomarkers from the 264 peak-containing m/z windows. In this study, we arbitrarily targeted selection of five m/z windows. The algorithm began with 100 particles where each particle consisted of 5 randomly selected m/z values from the 264 windows (i.e., n = 5, N = 100, and L = 264). A linear SVM classifier was built for each particle via the training dataset. The prediction power of each particle (five m/z windows) was evaluated by measuring the performance of the SVM classifier in distinguishing the two classes through the *k*-fold cross validation and bootstrapping methods. We used *k*=10 for this study. The most-fit particles contributed to the next generation of 100 candidate particles. This process

continued until the performance of the SVM classifier converged or a pre-specified number of iterations was reached. The algorithm was repeated 600 times, of which about half of the runs used the 10-fold cross-validation method and the other half used the bootstrapping method. Fig. 7 depicts the percentage of occurrence of m/z windows selected by the PSO-SVM. Note that the m/z windows are sorted in decreasing order of frequency and only the first 50 m/z windows are shown in the figure. Fig. 7 suggests that the first seven m/z windows are more frequently selected. Our TOF/TOF sequencing indicated that the first and the seventh m/z windows share the same sequence except for one amino acid. Thus, our subsequent analysis considered only the first six m/z windows. These six m/z windows yielded 100% sensitivity and 91% specificity in distinguishing liver cancer patients from healthy individuals in the testing dataset. Fig. 8 shows the box plot for the six m/z windows identified by the PSO-SVM algorithm. As shown in the figure, each of the six m/z windows is statistically significant candidate biomarkers.

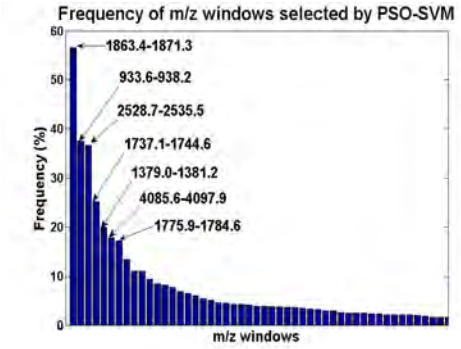


Fig. 7. Frequency of occurrence of m/z windows in 600 PSO-SVM runs for preprocessed spectra sorted in decreasing order of frequency (only the first 50 m/z windows are shown).

To examine the effect of data preprocessing on biomarker selection and sample classification, we performed biomarker selection using spectra that were binned and normalized, but not baseline corrected. 292 m/z windows were found from these spectra using our peak detection and alignment methods described before. The increase in the number of m/z windows is attributed to features that were not baseline corrected. The PSO-SVM algorithm was run 200 times with 100 particles to select 5 m/z windows out of 292 (i.e. n = 5, N = 100, and L = 292). The resulting frequency plot (Fig. 9) provided 5 biomarkers, of which the top 3 were very close to those found in the previous experiment. These 5 candidate biomarkers yielded 89% sensitivity and 86% specificity. This is significantly less than the prediction performance obtained when baseline correction was used in data preprocessing. To perform a fair comparison with the previous experiment, we tested the first six m/z windows from Fig. 9. However, the addition of the sixth m/z window did not improve the prediction accuracy. This shows that baseline correction has an impact in selecting biomarkers that provide improved sample classification.

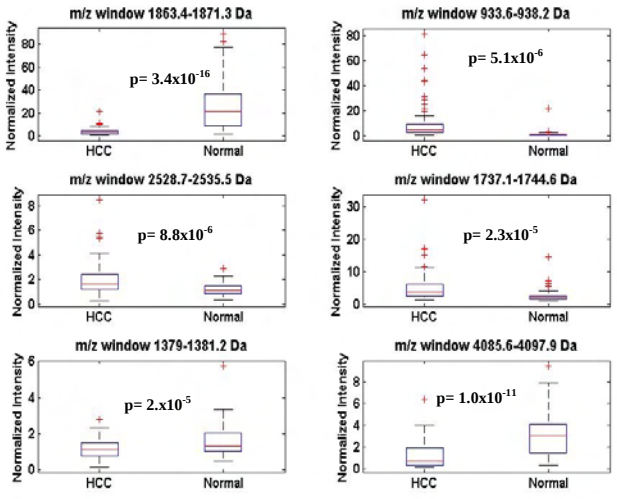


Fig. 8. Boxplots for the six m/z windows identified by the PSO-SVM. The boxplots show the distribution of each m/z window for HCC cases and normal using in both training and testing datasets combined.

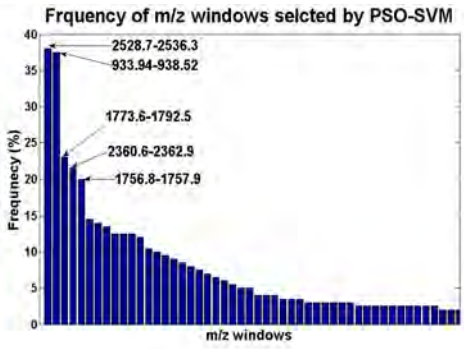


Fig. 9. Frequency of occurrence of m/z windows in 200 PSO-SVM runs for non-baseline corrected spectra, sorted in decreasing order of frequency (only the first 50 m/z windows are shown).

F. Sample Classification

We applied three classification algorithms, k-nearest neighbor (KNN), linear discriminant analysis (LDA), and SVMs to build classifiers. For comparison, we used three sets of features as inputs to the classifiers: all m/z bins, all m/z windows, and the six m/z windows selected by the PSO-SVM algorithm. Table 1 shows the sensitivity and specificity of the three classifiers in distinguishing HCC patients from healthy individuals in the testing dataset. Over all, the classifiers that used the six m/z windows performed better than those that used all m/z bins and m/z windows.

TABLE 1
PREDICTION ACCURACY OF THREE CLASSIFIERS ON THE TESTING DATASET.

Classification Methods	23,846 m/z bins		264 m/z windows		6 m/z windows	
	Sen.	Spec.	Sen.	Spec.	Sen.	Spec.
KNN (K=3)	96	77	96	73	93	91
LDA	89	91	89	95	98	92
SVM	93	91	93	86	100	91

III. CONCLUSIONS

This paper presents computational methods for preprocessing of mass spectral data, biomarker selection, and sample classification. Together, PSO and SVM are applied to identify candidate biomarkers from preprocessed MALDI-TOF spectra of enriched serum. The biomarkers distinguish cancer patients from non-cancer controls with high sensitivity and specificity. The PSO is used here to select a parsimonious subset from a large set of features. Since the particles contain discrete information only, we are currently investigating discrete methods such as ant colony optimization.

ACKNOWLEDGMENT

This work was supported in part by U.S. Army Medical Research and Material Command, Prostate Cancer Research Program grant DAMD17-02-1-0057 and American Cancer Society grant CRTG-02-245-01-CCE awarded to RG. We thank J. Jelinek for the programming of the binning algorithm. We thank Drs. F. Seillier-Moiseiwitsch and A. Wang for their helpful suggestions and discussions.

REFERENCES

[1] E. F. Petricoin, A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta, "Use of proteomic patterns in serum to identify ovarian cancer," *Lancet*, vol. 359, pp. 572-7, 2002.

[2] K. A. Baggerly, J. S. Morris, and K. R. Coombes, "Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments," *Bioinformatics*, vol. 20, pp. 777-85, 2004.

[3] E. P. Diamandis, "Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations," *Mol Cell Proteomics*, vol. 3, pp. 367-78, 2004.

[4] D. F. Ransohoff, "Bias as a threat to the validity of cancer molecular-marker research," *Nat Rev Cancer*, vol. 5, pp. 142-9, 2005.

[5] T. P. Conrads, V. A. Fusaro, S. Ross, D. Johann, V. Rajapakse, B. A. Hitt, S. M. Steinberg, E. C. Kohn, D. A. Fishman, G. Whitely, J. C. Barrett, L. A. Liotta, E. F. Petricoin, 3rd, and T. D. Veenstra, "High-resolution serum proteomic features for ovarian cancer detection," *Endocr Relat Cancer*, vol. 11, pp. 163-78, 2004.

[6] V. Paradis, F. Degos, D. Dargere, N. Pham, J. Belghiti, C. Degott, J. L. Janeau, A. Bezeaud, D. Delforge, M. Cubizolles, I. Laurendeau, and P. Bedossa, "Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases," *Hepatology*, vol. 41, pp. 40-7, 2005.

[7] Z. Zhang, R. C. Bast, Jr., Y. Yu, J. Li, L. J. Sokoll, A. J. Rai, J. M. Rosenzweig, B. Cameron, Y. Y. Wang, X. Y. Meng, A. Berchuck, C. Van Haaften-Day, N. F. Hacker, H. W. de Bruijn, A. G. van der Zee, I. J. Jacobs, E. T. Fung, and D. W. Chan, "Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer," *Cancer Res*, vol. 64, pp. 5882-90, 2004.

[8] E. F. Petricoin, 3rd, D. K. Ornstein, C. P. Paweletz, A. Ardekani, P. S. Hackett, B. A. Hitt, A. Velassco, C. Trucco, L. Wiegand, K. Wood, C. B. Simone, P. J. Levine, W. M. Linehan, M. R. Emmert-Buck, S. M. Steinberg, E. C. Kohn, and L. A. Liotta, "Serum proteomic patterns for detection of prostate cancer," *J Natl Cancer Inst*, vol. 94, pp. 1576-8, 2002.

[9] D. I. Malyarenko, W. E. Cooke, B. L. Adam, G. Malik, H. Chen, E. R. Tracy, M. W. Trosset, M. Sasinowski, O. J. Semmes, and D. M. Manos, "Enhancement of sensitivity and resolution of surface-enhanced laser desorption/ionization time-of-flight mass spectrometric records for serum peptides using time-series analysis techniques," *Clin Chem*, vol. 51, pp. 65-74, 2005.

[10] Y. Yasui, M. Pepe, M. L. Thompson, B. L. Adam, G. L. Wright, Jr., Y. Qu, J. D. Potter, M. Winget, M. Thornquist, and Z. Feng, "A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection," *Biostatistics*, vol. 4, pp. 449-63, 2003.

[11] A. C. Sauve, T. P. Speed, and "Normalization, baseline correction and alignment of high-throughput mass spectrometry data " *Proceedings of the Genomic Signal Processing and Statistics workshop, Baltimore, MD, USA.*, May 26-27, 2004.

[12] O. J. Semmes, Z. Feng, B. L. Adam, L. L. Banez, W. L. Bigbee, D. Campos, L. H. Cazares, D. W. Chan, W. E. Grizzle, E. Izbicka, J. Kagan, G. Malik, D. McLerran, J. W. Moul, A. Partin, P. Prasanna, J. Rosenzweig, L. J. Sokoll, S. Srivastava, S. Srivastava, I. Thompson, M. J. Welsh, N. White, M. Winget, Y. Yasui, Z. Zhang, and L. Zhu, "Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility," *Clin Chem*, vol. 51, pp. 102-12, 2005.

[13] L. Pusztai, B. W. Gregory, K. A. Baggerly, B. Peng, J. Koomen, H. M. Kuerer, F. J. Esteva, W. F. Symmans, P. Wagner, G. N. Hortobagyi, C. Laronga, O. J. Semmes, G. L. Wright, Jr., R. R. Drake, and A. Vlahou, "Pharmacoproteomic analysis of prechemotherapy and postchemotherapy plasma samples from patients receiving neoadjuvant or adjuvant chemotherapy for breast carcinoma," *Cancer*, vol. 100, pp. 1814-22, 2004.

[14] K. R. Coombes, S. Tsavachidis, J. S. Morris, K. A. Baggerly, M. C. Hung, and H. M. Kuerer, "Improved peak detection and quantification of mass spec-trometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform," The University of Texas M.D. Anderson Cancer Center, Technical Report UTMDABTR-001-04, 2004.

[15] E. T. Fung and C. Enderwick, "ProteinChip clinical proteomics: computational challenges and solutions," *Biotechniques*, vol. Suppl, pp. 34-8, 40-1, 2002.

[16] K. A. Baggerly, J. S. Morris, J. Wang, D. Gold, L. C. Xiao, and K. R. Coombes, "A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples," *Proteomics*, vol. 3, pp. 1667-72, 2003.

[17] W. Zhu, X. Wang, Y. Ma, M. Rao, J. Glimm, and J. S. Kovach, "Detection of cancer-specific markers amid massive mass spectral data," *Proc Natl Acad Sci U S A*, vol. 100, pp. 14666-71, 2003.

[18] J. Koopmann, Z. Zhang, N. White, J. Rosenzweig, N. Fedarko, S. Jagannath, M. I. Canto, C. J. Yeo, D. W. Chan, and M. Goggins, "Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry," *Clin Cancer Res*, vol. 10, pp. 860-8, 2004.

[19] H. Ressom, R. S. Varghese, D. Saha, E. Orvisky, L. Goldman, E. F. Petricoin, T. P. Conrads, T. D. Veenstra, M. Abdel-Hamid, C. A. Loffredo, and R. Goldman, "Particle swarm optimization for analysis of mass spectral serum profiles," *Proceedings of Genetic and Evolutionary Computation Conference (GECCO)*, vol. 1 pp. 431-438, 2005.

[20] H. W. Ressom, R. S. Varghese, M. Abdel-Hamid, S. Abdel-Latif Eissa, D. Saha, L. Goldman, E. F. Petricoin, T. P. Conrads, T. D. Veenstra, C. A. Loffredo, and R. Goldman, "Analysis of mass spectral serum profiles for biomarker selection," *Bioinformatics*, *in press*, 2005.

[21] R. S. Tirumalai, K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads, and T. D. Veenstra, "Characterization of the low molecular weight human serum proteome," *Mol Cell Proteomics*, vol. 2, pp. 1096-103, 2003.

[22] X. Zhang, S. M. Leung, C. R. Morris, and M. K. Shigenaga, "Evaluation of a novel, integrated approach using functionalized magnetic beads, bench-top MALDI-TOF-MS with prestructured sample supports, and pattern recognition software for profiling potential biomarkers in human plasma," *J Biomol Tech*, vol. 15, pp. 167-75, 2004.

[23] E. Orvisky, S. K. Drake, B. M. Martin, M. Abdel-Hamid, H. W. Ressom, R. S. Varghese, D. Saha, G. L. Hortin, C. A. Loffredo, and R. Goldman, "Enrichment of low molecular weight fraction of serum for mass spectrometric analysis of peptides associated with hepatocellular carcinoma," *Submitted to Proteomics*, 2005.

[24] S. Ezzat, M. Abdel-Hamid, S. Abdel-Latif Eissa, N. Mokhtar, N. A. Labib, L. El-Ghorory, N. N. Mikhail, A. Abdel-Hamid, T. Hifnawy, G. T. Strickland, and C. A. Loffredo, "Associations of pesticides, HCV, HBV, and hepatocellular carcinoma in Egypt," *Int J Hygiene Env Health*, *in press*, 2005.

[25] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B-splines and penalties," *Statist. Sci.*, vol. 11(2), pp. 89–121, 1996.

Ant Colony Optimization for Biomarker Identification from MALDI-TOF Mass Spectra

Habtom W. Ressom, Rency S. Varghese, Eduard Orvisky, Steven K. Drake, Glen L. Hortin, Mohamed M. Abdel-Hamid, Christopher A. Loffredo, and Radoslav Goldman

***Abstract*—We present a novel method that combines ant colony optimization with support vector machines (ACO-SVM) to select candidate biomarkers from MALDI-TOF serum profiles of hepatocellular carcinoma (HCC) patients and matched controls. The method identified relevant mass points that achieve high sensitivity and specificity in distinguishing HCC patients from healthy individuals. The results indicate that the MALDI-TOF technology could provide the means to discover novel biomarkers for HCC.**

I. INTRODUCTION

Analysis of peptides by MALDI-TOF mass spectrometry (MS) is an emerging technology for biomarker discovery. The method has a great potential to identify a panel of biomarkers relevant for early diagnosis of complex diseases such as cancer. Several laboratories have demonstrated the feasibility of selecting peptides in MALDI-TOF spectra for disease classification [1-4].

In our previous work [2, 5], we introduced a computational method that combines particle swarm optimization (PSO) with support vector machines (SVMs) for optimal selection of m/z values from SELDI-QqTOF and MALDI-TOF spectra. A limitation of the PSO algorithm is that it is not tailored for discrete optimization. We used PSO to search for discrete locations in high dimensional space by rounding the positions of the particles to the closest discrete location. In this paper, we present an alternative swarm intelligence-based approach known as ant colony optimization (ACO) that is particularly suitable for discrete optimization. We combined ACO with SVMs to identify the most relevant features (mass points). The algorithm lists these features in the order of their significance in predicting disease state. This will help prioritize candidate protein markers and panels for validation, which leads to assay development applicable to clinical settings.

The paper is organized as follows. Section II introduces our proposed ACO-SVM algorithm. Section III presents samples used in this study, sample preparation methods used to generate mass spectra, data preprocessing methods applied, and biomarkers identified by the ACO-SVM algorithm. Section IV concludes the paper.

II. ACO-SVM

Defined by Dorigo et al. [6], ACO studies artificial systems that take inspiration from the behavior of real ant colonies. The basic idea of ACO is that a large number of simple artificial agents are able to build good solutions to solve hard combinatorial optimization problems via low-level based communications. Real ants cooperate in their search for food by depositing chemical traces (pheromones) on the ground. Artificial ants cooperate by using a common memory that corresponds to the pheromone deposited by real ants. The artificial pheromone is accumulated at run-time through a learning mechanism. Artificial ants are implemented as parallel processes whose role is to build problem solutions using a constructive procedure driven by a combination of artificial pheromone and a heuristic function to evaluate successive constructive steps.

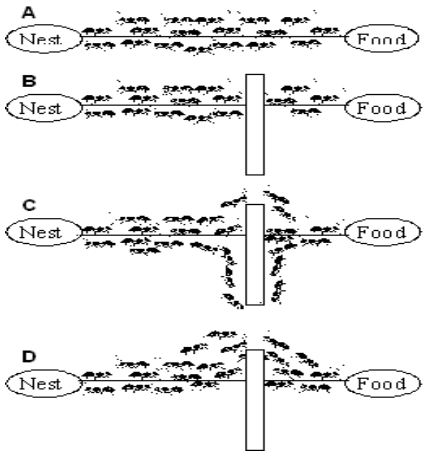


Fig. 1. Pheromone trail following of ants.

Figure 1 illustrates how ants select the shortest trail to fetch their food. The top figure (A) shows a single trail that all ants use to bring food to their nest. An obstacle is placed preventing ants to directly access the food (B). Initially, there is an equal chance for ants to take one of the two trails (note that the upper trail is shorter than the lower trail) (C). Later, ants choose to take the shorter trail (D) as those who used this trail come back to the nest faster than the others that use the second trail. As a result, more and more

This work was supported in part by U.S. Army Medical Research and Material Command, Prostate Cancer Research Program grant DAMD17-02-1-0057 and American Cancer Society grant CRTG-02-245-01-CCE awarded to RG.

H. W. Ressom, R. S. Varghese, E. Orvisky, C. A. Loffredo, and R. Goldman are with Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC (e-mail: hwr@georgetown.edu)

S. K. Drake and G. L. Hortin are with the Clinical Chemistry Service, Department of Laboratory Medicine, NIH, Bethesda, MD.

M.M. Abdel-Hamid is with the Viral Hepatitis Research Laboratory, NHTMRI, Cairo, Egypt.

pheromones will be deposited in the shorter trail over time thereby attracting the ants to use this trail.

We propose to use ACO for feature selection. To accomplish this, we use the probability function below:

$$P_i(t) = \frac{(\tau_i(t))^{\alpha} \eta_i^{\beta}}{\sum_i (\tau_i(t))^{\alpha} \eta_i^{\beta}}$$

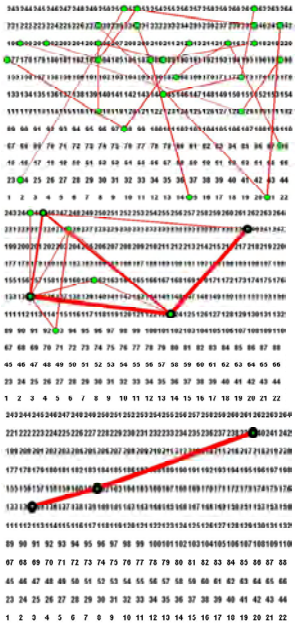
where $\tau_i(t)$ is the amount of pheromone trail for feature f_i (m/z window) at time t; η is a priori available information such as t-statistic or signal to noise ratio (SNR) for each feature; α and β are parameters which determine the relative influence of pheromone trail and a priori heuristic information, respectively.

At t=0, $\tau_i(t)$ is set to a constant for all features, allowing each feature to have equal probability of being selected. Thus, in the first iteration, ants choose randomly n distinct features (a trail) from L features. Let S_j be the jth ant consisting of n distinct features. Depending on the performance of S_j, the amount of pheromone trail for S_j will be updated. The performance function here is evaluated on the basis of disease state prediction capability of each S_j. We use the features in S_j to build an SVM classifier and estimate the prediction accuracy through the cross-validation method. The amount of pheromone trail for each feature in S_j is updated in proportion to prediction accuracy:

$$\tau_i(t+1) = \rho.\tau_i(t) + \Delta\tau_i(t)$$

where ρ is a constant between 0 and 1, representing the evaporation of pheromone trails. $\Delta\tau_i(t)$ is an amount proportional to the prediction accuracy achieved by S_j. $\Delta\tau_i(t)$ is set to zero, if $f_i \notin S_j$ at time t. This update is made for all M ants ($S_1,\ldots,S_M$). Note that at t=0, $\Delta\tau_i(t)$ is set zero for all features. The updating rule allows trails that yield good prediction to have their amount of pheromone trail increased, while others will evaporate. As the algorithm progresses, features with larger amounts of pheromone trails influence the probability function to lead the ants towards them.

To illustrate the ACO-SVM algorithm described above, we applied it to select three features from L=264. We used the SNR method proposed by Golub et al. (1999) as a priori heuristic information (η), α=β=1, and ρ=0.9. We define a feature as a location in the search space. Note that the dimension of the search space and the order of the features in the search space will not play a role, because the objective here is to maximize prediction accuracy, not distance between points. We placed the 264 features in a two-dimensional space where each location represents the labeled feature. M=10 ants were used to select n = three features. Initially, each ant chooses randomly three features (Fig. 2, top figure). The features selected are shown by the trails with three connected circles that lie on the selected features. At the 100th iteration, ants seem to favor some trails (middle figure). At the 284th iteration, all ants converged to one trail that goes through features 135, 162, and 240 (bottom figure). The prediction accuracy (found using the cross-validation method) improved from 79% at the 1st iteration to 91% at the 284th.



1st iteration			
Selected Sets			Accuracy %
118	189	206	79
231	188	15	78
240	24	252	75
.	.	.	.
.	.	.	.
262	173	145	55
168	88	216	50

100th iteration			
Selected Sets			Accuracy %
240	135	245	89
135	226	93	89
162	135	124	88
.	.	.	.
.	.	.	.
135	124	226	86
246	135	124	84

284th iteration			
Selected Sets			Accuracy %
240	162	135	91
240	162	135	91
240	162	135	91
.	.	.	.
.	.	.	.
240	162	135	91
240	162	135	91

Fig. 2. Pheromone trails for 10 ants at the 1st iteration (top), 100th iteration (middle) and 284th iteration (bottom).

III. Mass Spectral Data Analysis

A. *Sample collection*

Sample collection and generation of mass spectra was described previously [1, 2, 5]. The study examined an epidemic of viral infections in Egypt, a country where viral infections and associated HCC presents a serious health problem. The management of the disease would benefit from identification of biomarkers related to this disease. Serum samples of HCC cases and controls were obtained from Egypt between 2000 and 2002. Controls were recruited among patients from the orthopedic fracture clinic at the Kasr El-Aini Hospital (Cairo, Egypt) and were frequency-matched to cancer cases by gender, rural versus urban residency, and age [7]. Blood samples were collected in red top vacutainer tubes by trained phlebotomist each day around 10am and processed within a few hours according to a standard protocol. Aliquots of sera for mass spectrometric analysis were frozen at -80°C immediately after collection until analysis; all measurements were performed on samples of second-time thawed serum.

B. *Sample preparation*

The serum samples were enriched by denaturing ultrafiltration and desalting on C8 magnetic beads (MB) as described previously [1]. The procedure disrupts protein-protein interactions [8] and allows an efficient recovery of a Low Molecular Weight (LMW) serum fraction starting with 25 µl of serum. Eluted peptides were mixed with a matrix solution (3 mg/ml α-cyano-4-hydroxycinaminic acid in 50% actonitrile with 0.1% trifluoracetic acid), spotted onto

AnchorChip target (Bruker Daltonics, Billerica, MA) and analyzed using an Ultraflex MALDI TOF/TOF mass spectrometer (Bruker Daltonics). Each spectrum was detected in linear positive mode and was externally calibrated using a standard mixture of peptides. Denaturing ultrafiltration enriches LMW fraction of serum and plasma by removal of proteins greater than 50 kDa including albumin [1]. The enrichment improves quality of the spectra in the LMW region and allows analysis of approximately 300 peptides as described previously [1].

C. Data Preprocessing

Sixty-two replicate spectra were used to examine the run-to-run reproducibility of MALDI-TOF MS. Each spectrum consisted of about 136,000 m/z values with the corresponding ion intensities over the mass range 0.9 to 10 kDa. The dimension of the spectra was reduced to 23,846 m/z bins. A bin size of 100 ppm was found adequate. The mean of the intensities within each interval was used as the protein expression variable in each bin [9]. We transformed each intensity value by computing the base-two logarithm and found the mean log intensity value and standard deviation. The CV of the log-transformed intensity values in the 62 reference spectra ranged between 4.1% and 22.9% with a mean value of 10.5%.

For the remaining study, we used 84 HCC and 80 normal spectra. We excluded 14 spectra through outlier screening on the basis of their deviation from the median ion current, median record count (number of mass points), and their alignment with pre-selected landmarks. The remaining 150 spectra were binned, baseline corrected, and normalized. The baseline of each binned spectrum was estimated by obtaining the minimum value within a shifting window size of 50 bins. Spline approximation was used to regress the varying baseline. The regressed baseline was subtracted from the spectrum yielding a baseline corrected spectrum. We normalized each spectrum by dividing it its total ion current.

From the 150 preprocessed spectra (78 from patients with HCC and 72 from normal), we randomly selected 50 HCC and 50 normal (training spectra) for biomarker selection. The remaining 28 HCC and 22 normal (testing spectra) were set aside for later evaluation of the performance of the selected biomarkers.

We performed the following analyses using the 100 training spectra: (1) scaled the spectra to an overall maximum intensity of 100; (2) selected m/z values with reasonable intensity level and discarded those that appeared as noise, which was accomplished by identifying m/z values at which the slope sign changed from positive to negative and reasonable intensity was measured; (3) combined peaks if they differed in location by at most 7 bins or at most 0.03% relative mass. The method found 264 windows in the training spectra. For each spectrum, the maximum intensity within each window was found, yielding a 264 x 100 data matrix.

D. Biomarker Selection

We used the training spectra described in the previous section for biomarker (m/z window) selection. The validity of each classifier built with the selected biomarkers is evaluated using the sensitivity and specificity of the SVM classifier in distinguishing patients from healthy subjects. SVM classifiers are built for various combinations of m/z windows until the prediction accuracy of the SVM classifier converges or the maximum number of iteration is reached. The prediction accuracy is estimated through the four-fold cross-validation method.

To avoid any potential bias that may be introduced by parameter choice, the ACO-SVM algorithm was run for various numbers of features (n=3, 5, and 7) and ants (M=25, 50, and 100) with $\alpha=\beta=\gamma=1$, and $\rho=0.6$. Each combination (n features and M ants) was run 30 times, i.e., a total of 270 runs. Each run consisted of a maximum of 500 iterations. Figure 3 depicts the frequency of occurrence of the m/z windows in 270 runs. The figure suggests that the first seven m/z windows are frequently selected. Our TOF/TOF sequencing indicated that the first and the third m/z windows share the same sequence except for one amino acid. Thus, only the remaining six m/z windows are used in our subsequent analyses.

We used the SVM classifier to classify the testing spectra. We binned, baseline corrected, and normalized the testing spectra in the same way as the training spectra. Note that the testing spectra were scaled based on the parameters used to scale the training spectra. Figure 4 depicts the ROC curves and area under the ROC (AUROC) for the five markers both separately and combined by SVM. This figure demonstrates the advantage of a panel of biomarkers in achieving high prediction capability (100% sensitivity and 91% specificity) in distinguishing HCC patients from healthy individuals in the testing dataset.

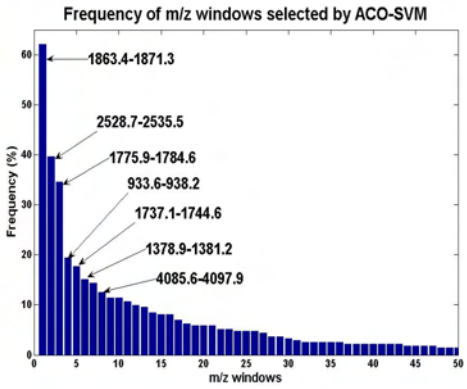


Fig. 3. Frequency of occurrence of m/z windows in 270 ACO-SVM runs sorted in decreasing order of frequency.

Finally, we compared three sets of features (the 23,846 m/z bins, 264 m/z windows, and the selected 6 m/z windows) in distinguishing HCC patients from healthy individuals using SVM classifiers. Note that each classifier was built using the training spectra and evaluated on the testing spectra. Figure

5 compares the ROC curves of the three SVM classifiers built using all bins, all m/z windows, and the six m/z windows. The figure shows that the AUROC for the SVM classifier with six m/z windows is larger than those that used all m/z bins or all m/z windows. Figure 6 shows the boxplots of the six m/z windows.

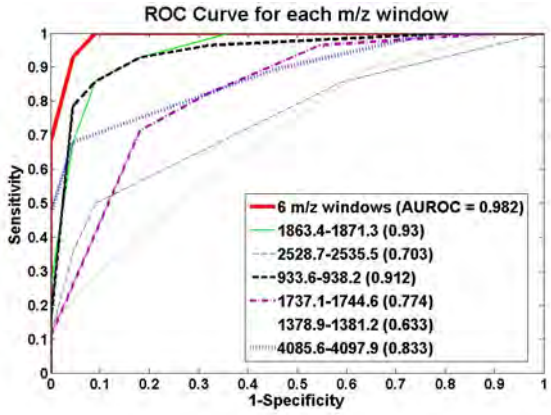


Fig. 4. ROC curves of each m/z window separately and all six combined. Note: the curves are based on testing spectra.

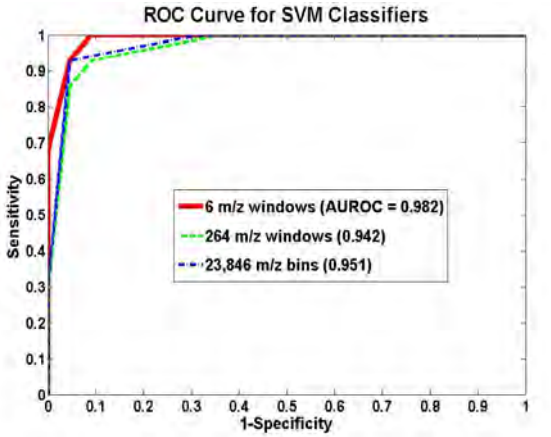


Fig. 5. ROC curves of three SVM classifiers (all bins, all m/z windows, and four m/z windows) on testing spectra.

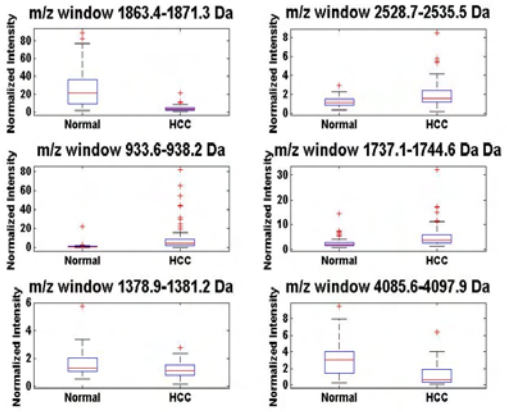


Fig. 6. Boxplots for the six m/z windows identified by the ACO-SVM (training and testing datasets combined).

IV. Conclusion

The high sensitivity and specificity achieved by the six candidate biomarkers indicate that MALDI-TOF technology, in conjunction with the proposed hybrid ACO-SVM algorithm could provide the means to discover novel biomarkers for HCC. The results also demonstrate the advantage of a panel of biomarkers in achieving high prediction capability.

Due to the initial trails which are determined randomly and the stochastic nature of the algorithm, every ACO-SVM run may not converge to the same trail in the search space. The frequency of occurrence of each m/z window in multiple runs allows us to estimate its relevance and the frequency response plot enables us to visually estimate the best number of m/z windows. Future work will focus on determining the frequency of occurrence of m/z windows that appear together (e.g. in pairs, triples, etc.) instead of combining the most frequent individual m/z windows via the frequency plot. The former will be useful to determine which m/z windows should be used together.

References

[1] E. Orvisky, S. K. Drake, B. M. Martin, H. Ressom, R. S. Varghese, D. Saha, G. L. Hortin, C. A. Loffredo, and R. Goldman, "Enrichment of low molecular weight fraction of serum for mass spectrometric analysis of peptides associated with hepatocellular carcinoma," *Proteomics In Press*, 2006.

[2] H. W. Ressom, R. S. Varghese, E. Orvisky, S. K. Drake, G. L. Hortin, M. Abdel-Hamid, C. A. Loffredo, and R. Goldman, "Analysis of MALDI-TOF serum profiles for biomarker selection and sample classification," *Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2005.

[3] J. Villanueva, D. R. Shaffer, J. Philip, C. A. Chaparro, H. Erdjument-Bromage, A. B. Olshen, M. Fleisher, H. Lilja, E. Brogi, J. Boyd, M. Sanchez-Carbayo, E. C. Holland, C. Cordon-Cardo, H. I. Scher, and P. Tempst, "Differential exoprotease activities confer tumor-specific serum peptidome patterns," *J Clin Invest*, vol. 116, pp. 271-84, 2006.

[4] X. Zhang, S. M. Leung, C. R. Morris, and M. K. Shigenaga, "Evaluation of a novel, integrated approach using functionalized magnetic beads, bench-top MALDI-TOF-MS with prestructured sample supports, and pattern recognition software for profiling potential biomarkers in human plasma," *J Biomol Tech*, vol. 15, pp. 167-75, 2004.

[5] H. W. Ressom, R. S. Varghese, M. Abdel-Hamid, S. Abdel-Latif Eissa, D. Saha, L. Goldman, E. F. Petricoin, T. P. Conrads, T. D. Veenstra, C. A. Loffredo, and R. Goldman, "Analysis of mass spectral serum profiles for biomarker selection," *Bioinformatics* vol. 21, pp. 4039-4045, 2005.

[6] M. Dorigo, G. Di Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," *Artif Life*, vol. 5, pp. 137-72, 1999.

[7] S. Ezzat, M. Abdel-Hamid, S. Abdel-Latif Eissa, N. Mokhtar, N. A. Labib, L. El-Ghorory, N. N. Mikhail, A. Abdel-Hamid, T. Hifnawy, G. T. Strickland, and C. A. Loffredo, "Associations of pesticides, HCV, HBV, and hepatocellular carcinoma in Egypt," *Int J Hygiene Env Health, in press*, 2005.

[8] R. S. Tirumalai, K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads, and T. D. Veenstra, "Characterization of the low molecular weight human serum proteome," *Mol Cell Proteomics*, vol. 2, pp. 1096-103, 2003.

[9] J. Villanueva, J. Philip, D. Entenberg, C. A. Chaparro, M. K. Tanwar, E. C. Holland, and P. Tempst, "Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry," *Anal Chem*, vol. 76, pp. 1560-70, 2004.

1. Optimization of Comet assay for quantification of DNA repair capacity in human whole blood. 97th Annual AACR Conference, Washington, DC, April 2006
Aleksandra Dakic, Allison Pollock, Michelle Ma, Daniel Saha, Sara Samie, Sherine Salem, Bozena Novotna, and **Radoslav Goldman**.

Purpose: We have begun a study of DNA repair capacity in peripheral blood lymphocytes. Comet assay was optimized for quantification of DNA repair capacity and applied to a pilot study of 10 healthy controls. Goal of the study is to evaluate whether lower DNA repair capacity correlates with higher risk of prostate cancer.
Methods: We optimized experimental conditions using Jurkat cells, primary lymphocytes isolated from peripheral blood, and whole blood cultures. Cells were embedded in agarose on microscopic slides and treated with ionizing radiation or a radiomimetic (bleomycin). Cells were incubated at 37°C for various time intervals, lysed, exposed to electric field, neutralized, fixed, and stained with ethidium bromide. Percentage of DNA in comet tails was used to quantify DNA damage and repair using a dedicated image analysis software (Loats Associates). Bleomycin (2µg/ml) induced migration of 80% DNA to the tail region. Most of the damage was repaired with a fast kinetic. Residual DNA in tail after 15 minutes of repair at 37^{o}C was on average 10%. Higher doses of bleomycin progressively increase the percentage of residual tail DNA. Staining with Anexin V showed that cultured lymphocytes repair their DNA efficiently, less than 15% of cells undergo apoptosis following exposure to bleomycin. Ionizing radiation induces lower migration of DNA to the tail region. The tail DNA following exposure to 5Gy of radiation is 30% and increases to 60% at 10Gy. Exposure of whole blood of 10 healthy controls to 9 Gy of radiation induced migration of 37-68% DNA to the tail region (mean 51%). The tail was repaired to 14-38% at 15 minutes (mean 29%) and to 7-28% at 45 minutes (mean 19%) after exposure following culture at 37^{o}C. Examination of frozen cells showed that the repair kinetic is slower, the variability of the repair measurement increases, and residual damage is higher compared to fresh cells. **Summary**: We optimized conditions of the Comet assay for screening of DNA repair capacity. Quantification of the rate and extent of DNA repair following radiation induced DNA damage may serve as a phenotypic measure of prostate cancer risk.

2. Yanming An, Mohamed Abdel-Hamid, Steve Drake, Lenka Goldman, Glen Hortin, Chris Loffredo, Eduard Orvisky, Habtom Ressom, Francoise Seillier-Moiseiwitsch, Rency Varghese, Antai Wang, and **Radoslav Goldman**. Analysis of serum peptides associated with hepatocellular carcinoma. 97th Annual AACR Conference, Washington, DC, April 2006

Purpose: Increasing incidence of hepatocellular carcinoma (HCC) in the US has been associated with hepatitis C (HCV) infections. We report a study of HCC in Egypt, a country with an epidemic of HCV and HCC. The goal of our study is to identify peptides in serum associated with HCC. These peptides will be used for early detection and improved classification of the disease. Methods and Results: We developed MALDI-TOF/TOF methods for analysis of serum peptides enriched by denaturing ultrafiltration. The methods were applied to a study of HCC. Serum samples were obtained in collaboration with NCI, Cairo, Egypt between 2000 and 2002. Controls were matched to

cases on gender, age, and residence. Analysis of TOF-MS spectra of 78 HCC cases and 72 controls in the 0.8-5 kDa mass range identified 264 peptides, a subset of which was identified by TOF/TOF sequencing. The abundance of 45 peptides was increased (34) or decreased (11) in patients with HCC. Using newly developed PSO-SVM computational methods, we selected 6 peptides that classify the disease with 100% sensitivity and 92% specificity in an independent set of 50 samples. Logistic regression analysis showed that the association of biomarker-candidates with HCC is not substantially altered by age, gender, residency, smoking, viral infections, and date of sample collection. Odds ratios (OR) of the six peptides significantly associated with HCC ranged from 1.3-3.3; the association remained significant in the adjusted models. The peptides were efficient at distinguishing stage I and II tumors and at distinguishing serum of HCC patients from serum of patients with cirrhosis (n=50). Conclusion: Using novel analytical methods, we identified six peptides that identify HCC with high prediction accuracy. A combination of six markers significantly improves the prediction accuracy of individual markers. These peptides may be useful in examining progression of chronic hepatitis C viral infection to malignancy.

3. An, Y; Ressom, HW; Varghese, SA; Goldman, L; Orvisky, E; Liao, J; Wang, A; Seillier-Moiseiwitsch, F; Drake, SK; Hortin, GL; Loffredo, CA and **Goldman, R**. MALDI-TOF analysis of serum peptides associated with hepatocellular carcinoma. AACR Special Conference, New Developments in the Epidemiology of Cancer Prognosis: Traditional and Molecular Predictors of Treatment Response and Survival. Charleston, South Carolina, January 2006.

Introduction: Increasing incidence of hepatocellular carcinoma (HCC) in the US has been associated with hepatitis C (HCV) viral infections. We report a study of HCC in Egypt, a country with an epidemic of HCV and HCC. The goal of our study is to identify serum peptides associated with HCC for early detection and improved classification of the disease. Methods: Serum samples were obtained in collaboration with NCI, Cairo, Egypt. Controls were recruited at the orthopedic fracture clinic and were matched to cases on gender, age, and residence (urban vs rural). We developed MALDI-TOF/TOF methods for analysis of serum peptides enriched by denaturing ultrafiltration and fractionation on magnetic beads. Analysis of TOF-MS spectra of 78 HCC cases and 72 controls in the 0.8-5 kDa mass range identified 264 peptides, a subset of which was identified by TOF/TOF sequencing. The abundance of 54 peptides increased (34) or decreased (20) significantly in patients with HCC based on randomized variance t-test. Using newly developed computational methods, we selected 6 peptides that classify the disease with 100% sensitivity and 92% specificity in an independent set of 50 samples. Logistic regression analysis showed that each of the six peptides is significantly associated with HCC. Odds ratios for three peptides increased in HCC range from 1.4 to 2.8; odds ratios of three peptides decreased in HCC range from 0.4 to 0.7. Association of the biomarker-candidates with HCC is not substantially altered by age, gender, viral infections, and date of sample collection. The peptides are efficient at distinguishing stage I and II tumors and at distinguishing serum of HCC patients from serum of patients with cirrhosis (n=50). Conclusion: Using novel analytical methods, we identified six peptides that identify HCC with high prediction accuracy. A combination of six markers significantly improves the prediction accuracy of individual markers. These peptides should be useful in examining

progression of chronic hepatitis C viral infection to malignancy. Development of a multiplex TOF-MS assay for quantification of the peptides is under way.

4. **Goldman, R**; An, Y; Liao, J; Orvisky, E; Ressom, HW; Varghese, SA; Goldman, L; Drake, SK; Hortin, GL; Loffredo, CA and Abdel-Hamid, M. MALDI-TOF analysis of serum peptides associated with hepatocellular carcinoma. ASPO, 30th Annual Meeting, Bethesda, MD, January 2006. Abstract of the oral presentation was published in *Cancer Epidemiol Biomarkers Prev* (2006) 15, 2, 409.

Purpose: Increasing incidence of hepatocellular carcinoma (HCC) in the US has been associated with hepatitis C (HCV) infections. We report a study of HCC in Egypt, a country with an epidemic of HCV and HCC. The goal of our study is to identify serum peptides associated with HCC for early detection and improved classification of the disease. Methods: Serum samples were obtained in collaboration with NCI, Cairo, Egypt. Controls were matched to cases on gender, age, and residence. We developed MALDI-TOF/TOF methods for analysis of serum peptides enriched by denaturing ultrafiltration. Analysis of TOF-MS spectra of 78 HCC cases and 72 controls in the 0.8-5 kDa mass range identified 264 peptides, a subset of which was identified by TOF/TOF sequencing. The abundance of 45 peptides was increased (34) or decreased (11) in patients with HCC. Using newly developed computational methods, we selected 6 peptides that classify the disease with 100% sensitivity and 92% specificity in an independent set of 50 samples. Logistic regression analysis showed that the association of biomarker-candidates with HCC is not substantially altered by age, gender, viral infections, and date of sample collection. Conclusion: Using novel analytical methods, we identified six peptides that identify HCC with high prediction accuracy. These peptides may be useful in examining progression of chronic hepatitis C viral infection to malignancy.

5. Ressom, H; Varghese, R; Dakic, A; Orvisky, E; Drake, SK; Hortin, GL; Abdel-Hamid, M; Loffredo, CA and **Goldman, R**. Analysis of MALDI-TOF Serum Profiles for Biomarker Selection and Sample Classification. American Association for the Study of Liver Diseases' (AASLD) Basic Research Single Topic Conference "Exploring the Functional Genomics and Proteomics of Liver in Health and Disease", Warrenton, VA, June 2005

Methods: We analyzed 150 MALDI-TOF mass spectra from 78 hepatocellular carcinoma patients (cases) and 72 healthy individuals (controls). Each spectrum consisted of about 134,500 mass-per-charge (m/z) values in the range between 919.7 and 9980.5 Da. The dimension of these spectra was reduced to 23,846 m/z bins via an algorithm that divided the m/z axis into intervals of desired length and calculated the mean intensity value within each bin. The baseline of each binned spectrum was estimated by obtaining the minimum value within a shifting window size of 50 bins. Spline approximation was used to regress the varying baseline. The regressed baseline was subtracted from the spectrum yielding a baseline corrected spectrum. Each binned and baseline-corrected spectrum was normalized by dividing it by its total ion current. The resulting spectra were split into training and testing datasets. The training dataset consisted of 100 samples (50 cases and 50 controls). The testing dataset had 50 samples (28 cases and 22 controls). The training spectra were rescaled so that the maximum intensity across all spectra is 100. We used

the training dataset for peak detection and peak alignment. A bin is identified as a peak if a change in the sign of the intensity's slope occurred. Those peaks with intensity below a pre-defined threshold-line were considered as noise and were discarded. To accommodate drifts introduced by the instrument, we aligned peaks by coalescing neighboring peaks within and across spectra into m/z windows. This approach resulted in 264 m/z windows. We applied an algorithm that combines support vector machines (SVMs) with particle swarm optimization (PSO) to select optimal biomarkers from the defined 264 m/z windows. The algorithm began with 100 particles where each particle consisted of 5 randomly selected m/z windows. A linear SVM classifier was built for each particle and its prediction power was evaluated through the k-fold cross validation and bootstrapping methods. SVM classifiers were built for various combinations of m/z windows guided by the PSO algorithm until a pre-specified maximum iteration number was reached. The algorithm was run 300 times and a frequency plot was used to determine the optimal biomarkers.
Results: Based on the frequency plot, we chose five m/z windows as biomarkers. These biomarkers yielded 100% sensitivity and 92% specificity in distinguishing liver cancer patients from healthy individuals in the testing dataset.
Conclusion: The results obtained in this study demonstrate the effectiveness of our data preprocessing and biomarker selection algorithms in identifying relevant biomarkers from complex spectra involving a large number of candidate biomarkers.

6. Orvisky, E; Ressom, H; Wang, A; Saha, D; Goldman, L; Petricoin, EF; Conrads, TP; Veenstra, TD; Liotta, LA; Drake, SK; Hortin, GL; Abdel-Hamid, M; Loffredo, CA and **Goldman, R**. Enrichment of low molecular weight (LMW) serum fraction for MALDI-TOF detection of hepatocellular carcinoma (HCC). 96th Annual AACR Conference, Anaheim, CA, March 2005

A challenging aspect of serum profiling is the dynamic range of serum proteins, i.e. interference of abundant proteins limiting number of peptides detected in the mass profile. We have previously reported serum profiling of HCC patients using a hybrid quadrupole time-of-flight (QSTAR, Applied Biosystems) mass spectrometer interfaced with a weak cation exchange (WCX) protein array (Ciphergen Biosystems). Comparison of the serum profiles of 70 cases and 70 controls by randomized variance t test ($p=0.0001$) with support vector machine (SVM) prediction algorithm led to the identification of a classifier that predicted disease status with 87% prediction accuracy in an analysis using leave-one-out cross-validation. The classifier, consisting of 136 m/z variables, achieved 89% prediction accuracy in an independent set of 78 cases and 87 controls. The prediction accuracy was improved to 93% when particle swarm optimization algorithm was used to select 15 m/z values as a classifier. When serum was desalted on magnetic beads and ultrafiltered under denaturing conditions, the mass spectra in the range of 1 to 10 kDa were greatly enriched and contained at least ten times more information. Denaturing conditions disrupts protein-protein interactions, which allows passage of the LMW components through the molecular weight cutoff membrane. Signal intensity of many peptides is enriched sufficiently for direct sequence analysis in TOF/TOF mode. Further reduction of sample complexity and increase of signal intensity is achieved by fractionation using strong cation exchange (SCX) and reverse phase C18

chromatography. Measurement of the enriched LMW serum proteome does not require SELDI interface and can be done on a MALDI platform of choice. We are currently measuring samples of HCC patients and matched controls on a MALDI TOF/TOF (Ultraflex, Bruker Daltonics). Further optimization of the enrichment/mass spectrometry and data analysis methods is under way. These data support the hypothesis that serum profiles can be used as a biomarker of disease status and should be evaluated in disease progression of chronic hepatitis C infection to malignancy.

7. Orvisky, E; Ressom, H; Saha, D; Goldman, L; Drake, SK; Hortin, GL; Abdel-Hamid, M; Loffredo, CA and **Goldman, R**. MALDI-TOF/TOF of enriched low molecular weight (LMW) serum fraction detects hepatocellular carcinoma (HCC). US Human Proteome Organization, 1st Annual Congress, Washington DC, March 2005

Dynamic range of serum proteins limits detection of low-abundance peptides. We have previously reported serum profiling of HCC patients using a hybrid quadrupole time-of-flight (QSTAR, Applied Biosystems) mass spectrometer interfaced with a weak cation exchange (WCX) protein array (Ciphergen Biosystems). Comparison of serum profiles of 70 cases and 70 controls by particle swarm optimization (PSO) with support vector machine (SVM) prediction algorithm selected 4 m/z values that classify HCC with 93% prediction accuracy in an independent set of 78 cases and 87 controls. When serum was enriched by denaturing ultrafiltration, the mass spectra in the range of 1 to 10 kDa were substantially enriched. Measurement of replicates of an enriched serum-standard on a MALDI TOF/TOF (Ultraflex, Bruker Daltonics) showed average coefficient of variance (CV) of 9% for 15 selected peaks. Analysis of HCC patients/controls (60/60) revealed significant differences between the groups using randomized variance t-test. An SVM classifier achieved greater than 90% prediction accuracy in a leave-one-out cross validation experiment. At least 20 well defined peaks were markedly different in cases and controls. Signal intensity of several differentially abundant m/z values was sufficient for a direct TOF/TOF sequence analysis. Further optimization of the enrichment/mass spectrometry and reduction of sample complexity by chromatographic fractionation is under way. These data support the hypothesis that analysis of enriched serum is a valuable method for biomarker discovery and should be applied to studies of disease progression of chronic hepatitis C infection to malignancy.

8. Orvisky, E; Ressom, H; Wang, A; Saha, D; Goldman, L; Drake, SK; Hortin, GL; Abdel-Hamid, M; Loffredo, CA and **Goldman, R**. Enrichment of low molecular weight (LMW) serum fraction for MALDI-TOF detection of hepatocellular carcinoma (HCC). Gordon Research Conference, New Frontiers in Cancer Detection and Diagnosis, Santa Ynez, CA, January 2005

A challenging aspect of serum profiling is the dynamic range of serum proteins, i.e. interference of abundant proteins limiting number of peptides detected in the mass profile. We have previously reported serum profiling of HCC patients using a hybrid quadrupole time-of-flight (QSTAR, Applied Biosystems) mass spectrometer interfaced with a weak cation exchange (WCX) protein array (Ciphergen Biosystems). Comparison of the serum profiles of 70 cases and 70 controls by particle swarm optimization with support vector machine (SVM) prediction algorithm was used to select 15 m/z values as a

classifier. The classifier achieved 93% prediction accuracy in an independent set of 78 cases and 87 controls. When serum was desalted on magnetic beads and ultrafiltered under denaturing conditions, the mass spectra in the range of 1 to 10 kDa were enriched and contained more information. Denaturing conditions disrupt protein-protein interactions, which allows isolation of the LMW components and elimination of albumin. Measurement of the enriched LMW serum proteome does not require SELDI interface and can be done on a MALDI platform of choice. Analysis of a pilot sample set of HCC patients and matched controls on a MALDI TOF/TOF (Ultraflex, Bruker Daltonics) revealed significant differences between the groups using randomized variance t-test with SVM classification and leave-one-out crossvalidation. Signal intensity of many peptides is enriched sufficiently for direct sequence analysis in TOF/TOF mode. Further reduction of sample complexity and increase of signal intensity is achieved by chromatographic fractionation of samples using strong cation exchange (SCX) and reverse phase C18 chromatography. These data support the hypothesis that serum profiles can be used as a biomarker of disease status and should be evaluated in disease progression of chronic hepatitis C infection to malignancy.